Published in final edited form as:

Biom J. 2017 November ; 59(6): 1352–1381. doi:10.1002/bimj.201600212.

Characterizing cross-subject spatial interaction patterns in functional magnetic resonance imaging studies: A two-stage point-process model

Adél Lee^{*,1}, Aila Särkkä², Tara M. Madhyastha³, and Thomas J. Grabowski⁴

¹Etosha Business and Research Consulting, PO Box 490003, Mount Berry GA 30149, USA

²Department of Mathematical Sciences, Chalmers University of Technology and University of Gothenburg, 41296 Gothenburg, Sweden

³Department of Radiology, University of Washington, 1959 NE Pacific Street, Box 357115, Seattle, WA, 98185, USA

⁴Department of Neurology and Department of Radiology, University of Washington, 1959 NE Pacific Street, Box 357115, Seattle, WA, 98185, USA

Abstract

We develop a two-stage spatial point process model that introduces new characterizations of activation patterns in multi-subject functional Magnetic Resonance Imaging (fMRI) studies. Conventionally multi-subject fMRI methods rely on combining information across subjects one voxel at a time in order to identify locations of peak activation in the brain. The two-stage model that we develop here addresses shortcomings of standard methods by explicitly modeling the spatial structure of functional signals and recognizing that corresponding cross-subject functional signals can be spatially misaligned. In our first stage analysis, we introduce a marked spatial point process model that captures the spatial features of the functional response and identifies a configuration of activation units for each subject. The locations of these activation units are used as input for the second stage model. The point process model of the second stage analysis is developed to characterize multi-subject activation patterns by estimating the strength of cross-subject interactions at different spatial ranges. The model uses spatial neighborhoods to account for the cross-subject spatial misalignment in corresponding functional units. We applied our methods to an fMRI study of 21 individuals who performed an attention test. We identified four brain regions that are involved in the test and found that our model results agree well with our understanding of how these regions engage with the tasks performed during the attention test. Our results highlighted that cross-subject interactions are stronger in brain areas that have a more specific function in performing the experimental tasks than in other areas.

*Corresponding author: etoshaconsulting@gmail.com.

Conflict of interest

The authors have declared no conflict of interest.

Keywords

Cross-subject interaction; Geyer saturation model; Misalignment in cross-subject functional signals; Multi-subject fMRI studies; Spatial point process model

1 Introduction

Functional Magnetic Resonance Imaging (fMRI) is a neuroimaging technique that helps us to understand the relationship between neural activation and mental activities, such as perception or motor processing tasks. During an fMRI session a series of three dimensional images of the brain is acquired at regular intervals over a period of about 3 to 5 minutes. The series of images acquired during a session reflects the blood oxygenation-level dependent (BOLD) responses in the brain.

To identify areas of activation in the brain, the series of brain volume images is typically considered to be a spatial collection of time-series recorded at each voxel location. Then, a linear model is fitted to each voxel time series. The resulting voxel-wise estimates of activation are then analysed spatially to identify brain regions that are involved in specific tasks.

Multi-subject fMRI studies are often used to gain insight into the relationship between task and neural activation in different populations of individuals. A question that has received much attention in the literature is how to combine information about activation across individuals. A standard approach in the fMRI literature has been to fit a linear mixed effects model to the multi-subject time-series data recorded at each voxel location (Beckmann *et al.*, 2003; Friston *et al.*, 2002a,b; Nichols and Holmes, 2002; Woolrich *et al.*, 2004; Worsley *et al.*, 2002).

One of the main shortcomings of this voxel-by-voxel (or ‘low-level’) analysis approach is that the spatial structure of the BOLD response is not accounted for. The BOLD signal arises from oxygenation changes in the capillaries surrounding the site of neural activity, as well as in larger veins that drain the activated tissue area. As a result the BOLD response is often spread over a neighbourhood of several voxels. These characteristics of the BOLD response are recognized in high-level imaging models that capture the location, shape and size of the BOLD response (Kim *et al.*, 2010; Kang *et al.*, 2011; Xu *et al.*, 2009; Hartvig, 2002; Weeda *et al.*, 2009).

Some authors considered specifically how high-level models can be used to combine information about activation across subjects. Kang *et al.* (2011) proposed a hierarchical spatial point process model that assumes that the foci (or locations) of activation that are identified for each subject are the offspring of a latent population process. The authors are specifically interested in how the population foci map to functional brain areas. Kim *et al.* (2010), on the other hand, used a nonparametric approach. The authors assume that there is a true underlying image template and that the individual subject patterns of activation are noisy realizations of the true template. Xu *et al.* (2009) developed a Bayesian hierarchical

mixture model approach. The authors were also interested in identifying population units of activation and in modeling the variation of the individual centers around population units.

In this paper, we also use high level statistical image analysis methods to identify the image elements— or ‘activation units’—of interest. These units summarize the location, height (magnitude) and size (spatial extent) of activation for each subject. Our approach is based on using a spatial marked point process model to identify the activation units of each subject (Baddeley and Lieshout, 1993; Al-Awadhi *et al.*, 2011; Hartvig, 2002).

Rather than using the activation units to draw inference on latent population foci, our focus will be on studying the cross-subject spatial interactions in these units. We achieve this by developing a two-stage model. In the first stage, we model the activation units for each subject as a marked spatial point-process. Then, in the second stage analysis, we develop a pairwise point-process model to study the cross-subject interactions in the locations of these units. We will use the terminology ‘first stage model’ and ‘second stage model’ to distinguish between the two components of the model. Even though both components of the model are essential to answering our scientific questions of interest, in this work our main focus is on the development and contributions of the second stage model.

The main focus of this paper will be to illustrate the development of our second stage model. The model that we develop is motivated by Geyer’s saturation model (Geyer, 1999). Our goal is to develop a parsimonious model that will give us insight into how likely it is that a subject has activation in a region of interest, given that other subjects in the group have activation in that region. By estimating the interaction parameter of our model, we can compare how consistent activation is in different brain regions or across different groups of individuals.

The second stage spatial point process model that we develop introduces novel features in the point-process literature. First, rather than studying the interactions between different types of points on the same image (Eckel *et al.*, 2009; Diggle *et al.*, 2006), we study the interactions between the points from subject-pairs. Thus, our replicates are pairs of subjects instead of individual subjects, locations, or images (Myllymäki *et al.*, 2012; Waller *et al.*, 2011; Schladitz *et al.*, 2003; Diggle *et al.*, 2000; Baddeley *et al.*, 1993). This introduces a complicated dependence structure between our replicates that needs to be accounted for.

The second feature that our study introduces is that cross-subject points that are the ‘same’ (*i.e.* they correspond to the same functional signal) may not occur at the same voxel location. In fMRI studies, anatomical landmark normalization is used to register the functional images of each subject to a standard brain template. The goal of spatial registration is to bring functionally homologous areas in the closest possible spatial alignment across the individuals. However, it has been highlighted in the literature that a fair amount of functional variability remains across subjects after registration (Frost and Goebel, 2012; Thirion *et al.*, 2010; Xu *et al.*, 2009; Thirion *et al.*, 2006; Brett *et al.*, 2002). As a result, our model needs to adjust for the fact that cross-subject points that correspond to the same functional signal are not necessarily aligned.

We use a Bayesian framework to specify our first and second stage models and do inference via Markov chain Monte Carlo.

Our paper is organized as follows. In Section 2 we describe the data set used in our analysis, and in Sections 3 and 4 we specify the first and second stage models, respectively. The details of our MCMC algorithm and model implementation are given in Appendices A, B, C. We fit our model to the fMRI data set in Section 5 and end with a discussion of our results in Section 6.

2 fMRI data

We use data from a multi-subject fMRI study conducted by the Integrated Brain Imaging Center at the University of Washington. This study includes a group of 21 control individuals (9 males) with an average age of 62 (standard deviation 10). The subjects performed an attention network test (Fan *et al.*, 2005). The test is designed to test three attentional networks, and it consists of a series of ‘target’ and ‘cue’ prompts that the subject has to respond to.

In this paper we consider activation corresponding to the target condition only. The target conditions consisted of five visually presented arrows (arranged horizontally in one line). The central arrow (or target) either pointed in the same direction as the four flanking arrows (congruent target) or in the opposite direction (incongruent target) as the four flanking arrows. The participant had to identify the direction of the target arrow by pressing a button with the left index finger if the central arrow pointed leftward or by pressing a button with the right index finger if the arrow pointed rightward. Response to the target activated widespread regions of the brain involved in attention, vision and motor control.

The characteristics of the attention network test, MRI acquisition and data pre-processing are described in detail by Madhyastha *et al.* (2015a). As part of pre-processing the brain volume data, the individual subjects’ images are registered to the standard Montreal Neurological Institute (MNI) space. This step allows us to make voxel-by-voxel comparisons across subjects. A linear model was fitted to the time-series at each voxel location using the brain imaging software, FSL’s (Jenkinson *et al.*, 2012) FMRI Expert Analysis Tool (FEAT) version 6.0. For this analysis, the time-series data, recorded during an fMRI session, is regressed against explanatory variables that describe the stimulus conditions and other covariates that account for known sources of noise in the data. Note that, for the purpose of our analysis, a separate linear model was fitted for each subject. (This can be contrasted with the standard method in the fMRI literature—using linear mixed effects models to combine information about activation across subjects.)

The linear model regression coefficients measure of the strength of the relationship between the observed data and the experimental conditions. For our analysis, we are interested in the coefficient estimate representing the target condition. This estimate is transformed into a z -statistic. This involves calculating the probability of observing a value less than or equal to the coefficient estimate under the relevant distribution. The calculated probability is then transformed into the corresponding z -statistic under the standard normal distribution. One z -

statistic is calculated for each voxel, and the resulting volumes of z -statistics (one volume per subject) are then analyzed to identify regions of activation.

We select an axial slice (MNI coordinate $Z = -2$) that intersects the inferior lateral occipital cortex, the frontal operculum and insula for our analysis. The target condition involves the attention, motor and premotor networks for the purpose of planning and executing the button press. The insula is part of the salience network that helps in network selection during this task. The lateral occipital cortex is engaged in the higher order visual processing of the target. Note that the dimension of the voxels in this dataset is $1 \times 1 \times 1 \text{ mm}^3$. We can view the slice as a grid with sides that are 1mm. The brain region of the selected slice is 192 voxels high and 159 voxels wide (that is, the maximum number of voxels in the y and x directions).

By selecting a slice, we extract 21 maps (sets of z -statistics) that are used as input for the first stage of our model. We only consider areas with positive z -values for this analysis. Therefore, voxels with z -values less than 0 are assigned a value of 0. We display the map of z -statistics for four of the subjects in our dataset in Figure 1. Note the variability in height and size of activation among the subjects.

3 Model specification: First stage model

As mentioned above, we develop a two-stage model in this paper. In the first stage model, we use the activation images (spatial maps of z -statistics) as input and identify the image elements of interest. These elements or ‘activation units’ capture the features of the BOLD responses that we expect to see in activated brain regions. In particular, we are looking for spatially extended areas that are smooth and show a high level of activation. The activation units describe such regions in terms of the location, height (strength), and size (spatial extent) of activation.

3.1 Likelihood

We denote the activation surface observed for subject s , $s = 1, \dots, S$, by the set of intensity measurements $\mathbf{y}_s = \{y_{s,v} | v \in V\}$. Here V is the two-dimensional grid on which the data is observed.

We model the observed activation surface for subject s as the sum of an underlying activation pattern and a noise term,

$$y_{s,v} = A_{s,v} + \eta_{s,v}, \quad \eta_{s,v} \sim \mathcal{N}(0, \tau^2),$$

for $s = 1, \dots, S$ and $v \in V$. We assume that the noise $\eta_{s,v}$ is independently distributed across the grid and the subjects.

The underlying activation surface $\mathbf{A}_s = \{A_{s,v} | v \in V\}$ for subject s is a deterministic function, given the configuration of marked points $\mathbf{x}_s^M = \{x_{s,1}^M, \dots, x_{s,n_s}^M\}$ for subject s with n_s activation

units. We refer to each marked point $x_{s,j}^M = (x_{s,j}, a_{s,j}, d_{s,j})$ as an *activation unit*. Each unit summarises the location x of activation on the grid V , and the height a and size d associated with it. To make the relationship between $A_{s,v}$ and x_s^M clear, $A_{s,v} = A(v; x_s^M)$ and

$$A(v; x_s^M) = h(v; x_{s,1}^M) + \dots + h(v; x_{s,n_s}^M),$$

for $v \in V$.

The function $h(v; x_{s,j}^M)$ is a Gaussian bell. The functional form of each bell is $h(v; (x, a, d)) = ad \times f(v)/\log 2$ where $f(v) = f(v; x, \Sigma)$ is the multivariate normal density function with mean x and covariance matrix Σ ,

$$\Sigma = \frac{d}{2\pi \log 2} I.$$

This specification of the covariance matrix, where I is a two-dimensional identity matrix, implies that we only consider circular bells. In this context, a is interpreted as the maximum height of the Gaussian bell, *i.e.* $h(x; (x, a, d)) = a$, and d is the area of the cross section of the Gaussian bell at half of the maximum height. The cross section of the bell is taken so that it is parallel to the level of the grid, implying that the cross section is a disk with area d .

3.2 Prior model for the configuration of activation units

We require a prior model that allows us to identify the underlying units of activation accurately. In this section we define a marked spatial point process prior that enables us to pick up the activation signal by enforcing a minimal set of constraints.

Location of activation—In the first stage model, we assume that the number of subjects S is fixed and that the number of activation units for each subject, n_s , is random. Here we assume that the marks (height and size) are independent of the locations, based on the assumption that prior knowledge on how the spatial extend and amplitude of the BOLD signals vary by specific brain region may not be available. In addition, we assume independence between the marks.

We define a hard-core prior on the location points of each subject,

$$p(x_{s,1}, \dots, x_{s,n_s}) = \begin{cases} \prod_{j=1}^{n_s} \beta(x_{s,j}) & \text{if } \|x_{s,j} - x_{s,j'}\| > R_W \forall j \neq j', \\ 0 & \text{otherwise,} \end{cases} \quad (1)$$

for $s = 1, \dots, S$.

This prior specification is motivated by our understanding of the BOLD response. The BOLD signal arises from oxygenation changes in the capillaries surrounding the site of

neural activity. In order to identify separate, non-overlapping areas of activation, we enforce a minimal distance, R_W , that the location points should be apart within subjects. We set R_W based on an exploratory analysis of the data as discussed in Section 3.3.

The first order term β can be modeled as a function of a covariate that reflects how the intensity of the point pattern is expected to change across the brain. We calculate the covariate to be equal to the smoothed average of the spatial maps y_s , $s = 1, \dots, S$.

Note that the prior for the location parameters is independent across subjects. This is an intentional choice: we want the locations of subject s to be determined by y_s only. An alternative specification is to allow for point location information to be shared across subjects (Bothma, 2010).

Height and size of activation units—As mentioned above, the marks are distributed independently from the locations of the activation centres and from each other. We also assume that the height and size parameters are independent within and across subjects. The prior on the height, p_a , is chosen so that lower values of a are penalised. We choose an inverse-gamma prior for a , $a \sim \text{Inv-gam}(\alpha_a, \beta_a)$, $a \in (0, C_a]$, with probability density function

$$p_a(a) \propto a^{-(\alpha_a + 1)} \exp(-1/(a\beta_a)). \quad (2)$$

This distribution gives us control over penalising smaller values of a through adjusting α_a and β_a . Also note that the prior is truncated to the interval $(0, C_a]$. This is only a practical constraint that is helpful when implementing the model. (We set C_a to exceed the maximum intensity, i.e. $C_a > \max \{y_v : v \in V\}$.)

We specify a uniform prior for the size, $d \sim \text{Unif}(d_{low}, d_{up})$, for $d \in [d_{low}, d_{up}]$. The interval $[d_{low}, d_{up}]$ is set to correspond to the spatial scale of the BOLD response.

3.3 Setting model parameters and hyperparameters

Here we describe the methods used to choose values for the model variance parameter τ^2 and the hyperparameters R_W , α_a , β_a , C_a , d_{low} , d_{up} .

- To specify the variance of the likelihood model, τ^2 , we recall that under the null hypothesis of no activation in the brain volume the map of z -statistics should behave like samples from the standard normal distribution. In order to identify areas where the z -statistics deviate from this behavior, we set $\tau^2 = 1$.
- To get insight into setting the hard-core distance R_W , we did an exploratory analysis of the activation surfaces of the 21 subjects. This was done by smoothing the activation surface of each subject and identifying the local maxima of the smoothed surfaces. (Smoothing was implemented by convolving each functional image with a fixed-width Gaussian filter.) We explored the nearest-neighbor distances between the local maxima and found that all nearest neighbor maxima were at least 5 voxels apart within subjects. The voxel

dimension for the data that we analyze in this paper is $1 \times 1 \times 1 \text{ mm}^3$. Based on this, we set the within-subject hardcore distance R_W to be equal to 5mm.

- To specify d_{low} and d_{up} we recall that the size d is the area of the disk that is the cross-section of an activation unit (Gaussian bell) at half its height. We want to constrain the size of this region so that neighboring activation units can overlap (based on our understanding that the BOLD signal often extends beyond the capillary network adjacent to the activated neurons). As a result, we let the maximum diameter of the disk be 2 times the hardcore distance R_W , i.e. d_{up} is set equal to $5^2 \pi \text{ mm}^2$. We set $d_{low} = 2.5^2 \pi \text{ mm}^2$, so that our model can detect sharp, localized peaks of activation. Then, the minimum diameter of the disk is 5mm.
- The height hyperparameters α_a and β_a should be chosen so that low values of a are penalized. We designed a simulation study and generated activation surfaces with activation units with heights of 1 or more and with a variety of sizes. We found that the setting $\alpha_a = 10$ and $\beta_a = 0.023$ allowed us to detect the true activation units with high accuracy.

3.4 Posterior distribution

The posterior density function of the first stage model is

$$p(\mathbf{x}^M | \mathbf{y}) \propto \begin{cases} \exp\left\{-\frac{1}{2\tau^2} \sum_{s=1}^S \sum_{v \in V} (y_{s,v} - A_{s,v})^2\right\} \times \text{if } \|x_{s,j} - x_{s,j'}\| > R_W \forall j \neq j' \\ \prod_{s=1}^S \prod_{j=1}^{n_s} \beta(x_{s,j}) p_a(a_{s,j}) p_d(d_{s,j}) & \\ 0 & \text{otherwise} \end{cases},$$

for $\mathbf{x}^M = \{\mathbf{x}_1^M, \dots, \mathbf{x}_S^M\}$, the marked point configuration for the set of S subjects.

4 Model specification: Second stage model

As input to the second stage model we use the location points $\mathbf{x}$ that correspond to the configuration of activation units $\mathbf{x}^M$ that maximized the posterior $p(\mathbf{x}^M | \mathbf{y})$. We consider the point pattern $\mathbf{x}$ to be a multitype process, with each subject a different type. The subscript s is used to distinguish between subjects (or types) and $\mathbf{x}_s = \{x_{s,1}, \dots, x_{s,n_s}\}$ denotes the set of location points for subject s .

Our goal with the second stage model is to develop a model that a) accounts for the cross-subject spatial misalignment in homologous functional areas and b) allows us to advance our understanding of how the cross-subject patterns of clustering vary across brain regions or among different groups of individuals. With these aims in mind we develop a pairwise interaction model. Pairwise interaction point process models are typically suitable for modeling regularity but not clustering. Various models have been proposed in the literature to overcome this limitation, e.g. the Geyer's saturation model (Geyer, 1999), area-interaction point processes (Baddeley and Van Lieshout, 1995), continuous random cluster Markov

point processes (Møller, 1999), and interacting neighbor processes (Grabarnik and Särkkä, 2001). We decided to use the Geyer's saturation model as a basis for our second stage model, as it is a parsimonious model that can also be extended to deal with multitype data.

The Geyer saturation process model for univariate homogeneous data has a density function of the form

$$p(\mathbf{x}) = \frac{1}{Z_{\beta, \gamma}} \beta^n \gamma^{s(\mathbf{x})} \quad (3)$$

where $Z_{\beta, \gamma}$ denotes the normalizing constant, β the intensity parameter, n is the number of points in $\mathbf{x}$, and γ is the interaction parameter. The function $s(\mathbf{x})$ is the sum of the thresholded number of neighbors,

$$s(\mathbf{x}) = \sum_{j=1}^n \min(c, \sum_{j' \neq j} \mathbf{1}[\|x_j - x_{j'}\| \leq R]) = \sum_{j=1}^n \min(c, N_{\mathbf{x}_{-j}}(x_j)).$$

Here $N_{\mathbf{x}_{-j}}(x_j)$ is the number of neighbors that the point x_j has in the set of points $\mathbf{x}_{-j} = \{x_1, \dots, x_{j-1}, x_{j+1}, \dots, x_n\}$. A saturation threshold, c , is specified for the number of neighbors. We define two points to be neighbors if they are less or equal to a distance of R apart. For $\gamma = 1$ the model simplifies to the Poisson process, for $\gamma < 1$ the density (3) models repulsion, and for $\gamma > 1$ clustering.

To study cross-subject patterns, we extend the model to a bivariate model that describes the interaction between pairs of subjects. For subject pair (s, s') , we define the density

$$p(\mathbf{x}_s, \mathbf{x}_{s'}) = \begin{cases} \frac{1}{Z_{\beta, \gamma_B}} \beta^{n_s + n_{s'}} \gamma_B^{s_B(\mathbf{x}_s, \mathbf{x}_{s'})} & \text{if } \|x_{s^*, j_1} - x_{s^*, j_2}\| > R_W \text{ } s^* \in \{s, s'\} \text{ and } \forall j_1 \neq j_2, \\ 0 & \text{otherwise.} \end{cases} \quad (4)$$

As before, Z_{β, γ_B} denotes the normalizing constant and β the common intensity parameter.

We assume a constant intensity across subjects. This assumption is discussed in Section 4.3.

Note that we enforce the same hard-core constraint as in model (1), that all pairs of within-subject points should be at least a distance of R_W apart.

Contrary to model (3), the interaction parameter γ_B now captures the strength of between-subject interactions and the function s_B counts the number of between-subject neighbors. The new neighborhood function, $N_{B\mathbf{x}_{s'}}(x_{s, j})$, counts the number of cross-subject points that

are within distance R_B apart. We assume that activation units in the same neighborhood reflects the same underlying functional response to task across subjects.

The total number of between-subject neighbors is given by the function

$$\begin{aligned} s_B(\mathbf{x}, \mathbf{x}_{s'}) &= \sum_{j=1}^{n_s} \min(c, \sum_{j'=1}^{n_{s'}} \mathbf{1}[\|\mathbf{x}_{s,j} - \mathbf{x}_{s',j'}\| \leq R_B]) + \sum_{j'=1}^{n_{s'}} \min(c, \sum_{j=1}^{n_s} \mathbf{1}[\|\mathbf{x}_{s',j'} - \mathbf{x}_{s,j}\| \leq R_B]) \\ &= \sum_{j=1}^{n_s} \min(c, N_{B\mathbf{x}_{s'}}(\mathbf{x}_{s,j})) + \sum_{j'=1}^{n_{s'}} \min(c, N_{B\mathbf{x}_s}(\mathbf{x}_{s',j'})). \end{aligned}$$

Here $N_{B\mathbf{x}_{s'}}(\mathbf{x}_{s,j})$ is the number of between-subject neighbors that the point $\mathbf{x}_{s,j}$ has in $\mathbf{x}_{s'} = \{\mathbf{x}_{s',1}, \dots, \mathbf{x}_{s',n_{s'}}\}$.

The above model assumes a common interaction parameter γ_B for all subject pairs (Diggle et al., 2000; Eckel et al., 2009). Illian and Hendrichsen (2010) and Bell and Grunwald (2004) have highlighted that the methodology for the analysis of replicated point patterns is quite limited. They use an approach based on generalized linear mixed models to relax the assumption of a common interaction parameter in replicated data. The authors show how maximum pseudolikelihood methods (Baddeley and Turner, 2000) can be used to obtain model estimates.

We also relax this assumption of a common interaction parameter, but use a fully Bayesian approach to specify the model and draw inference. To simplify notation, we denote our subject pairs as $\mathbf{x}_{\{i\}} = (\mathbf{x}_{s_i}, \mathbf{x}_{s'_i})$, and we fit our model to a set of m such subject pairs: $\mathcal{S}_m = \{\mathbf{x}_{\{1\}}, \dots, \mathbf{x}_{\{m\}}\}$. Then,

$$p(\mathbf{x}_{\{i\}}) = \begin{cases} \frac{1}{Z_{\beta, \gamma_i}} \beta^{n_{s_i} + n_{s'_i}} + \gamma_i^{s_B(\mathbf{x}_{\{i\}})} & \text{if } \|\mathbf{x}_{s,j_1} - \mathbf{x}_{s,j_2}\| > R_W \text{ for } s \in \{s_i, s'_i\} \text{ and } \forall j_1 \neq j_2, \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

for $\mathbf{x}_{\{i\}} = (\mathbf{x}_{s_i}, \mathbf{x}_{s'_i})$ and $\mathbf{x}_{\{i\}} \in \mathcal{S}_m$. We assign a lognormal prior distribution to the m interaction parameters

$$\gamma_i \sim \text{logn}(\mu_\gamma, 1/\tau_\gamma^2). \quad (6)$$

and hyperprior distributions to the mean μ_γ and precision τ_γ^2 :

$$\mu_{\gamma} \sim N(\mu_0, 1/\tau_0^2)$$

and

$$\tau_{\gamma}^2 \sim \text{gamma}(\alpha_0, \beta_0).$$

4.1 Independent subjects pairs

The above model specification in (5) and (6) does not recognize that the γ_i 's that share a subject are dependent. As an example, if subject s has a sparse point pattern relative to other subjects, then γ_i will be lower for all subjects pairs that include subject s . On the other hand, the γ_i 's will be higher for all subject pairs that include subject s' that has a very dense pattern. To address this issue one could adjust the prior for the γ_i 's to reflect the dependence between subsets of γ_i that share a subject. However, the specification of such a prior can get very complex for studies with a relatively large number of subjects. (Our study includes 21 subjects and therefore 210 distinct pairs of subjects.)

We consider a simplified approach and fit our model to 'independent' subject pairs. These are subject pairs that do not share a subject. We define $\mathcal{S}_m^*$ to be a set of independent subject pairs: $\mathcal{S}_m^* = \{(x_{s_1}, x_{s'_1}), (x_{s_2}, x_{s'_2}), \dots, (x_{s_m}, x_{s'_m})\}$ such that $\{s_1, s'_1, s_2, s'_2, \dots, s_m, s'_m\}$ is a set of unique integers (no duplicate subjects). In addition, we define $\mathcal{S}_m^* = \{(s_1, s'_1), (s_2, s'_2), \dots, (s_m, s'_m)\}$ to be the set of corresponding subject indices.

4.2 Range of interaction

We use studies on inter-subject variability in activation locations to inform us on setting the range of interaction R_B . Frost and Goebel (2012) studied the mean distance from the peak location of activation for each subject to the group mean location, after spatial normalization based on anatomical landmarks. The authors calculated this mean distance in several functional regions of the brain and found that the distance is mostly less than 10mm. Thirion *et al.* (2006) dealt with the shortcomings of spatial registration in multi-subject data by developing a method that finds population level activation prototypes. They used a cut-off distance of 10mm for a voxel to be matched to a prototype. Based on these results and exploratory analysis results of our data, we set $R_B = 10\text{mm}$. Our goodness-of-fit results for this setting, however, indicated that our assumption that the strength of interaction γ is constant over this range is too limiting. We adjusted our model to allow for two interaction parameters—one corresponding to shorter-range interactions and the other to longer-range interactions. Then,

$$p(\mathbf{x}_{\{i\}}) = \quad (7)$$

$$\begin{cases} \frac{1}{Z_{\beta, \gamma_1, \gamma_2}} \beta^{n_{s_i} + n_{s'_i}} \gamma_1^{s_1(\mathbf{x}_{\{i\}})} \gamma_2^{s_2(\mathbf{x}_{\{i\}})} & \text{if } \|x_{s, j_1} - x_{s, j_2}\| > R_W \text{ for } s \in \{s_i, s'_i\} \text{ and } j_1 \neq j_2, \\ 0 & \text{otherwise} \end{cases}$$

for $\mathbf{x}_{\{i\}} = (\mathbf{x}_{s_i}, \mathbf{x}_{s'_i})$ and $\mathbf{x}_{\{i\}} \in \mathcal{S}_m^*$. Note that where we previously used the subscript B to indicate between-subject interactions, we now drop this subscript for notational simplicity and use 1 and 2 instead to distinguish between shorter-range and longer-range inter-subject interactions, respectively.

The functions s_1 and s_2 count the number of between-subject neighbors over shorter and longer ranges, respectively:

$$s_1(\mathbf{x}_{\{i\}}) = \sum_{j=1}^{n_s} \min(c, N_{\mathbf{x}_{s'}}^1(x_{s, j})) + \sum_{j'=1}^{n_{s'}} \min(c, N_{\mathbf{x}_s}^1(x_{s', j'}))$$

and

$$s_2(\mathbf{x}_{\{i\}}) = \sum_{j=1}^{n_s} \min(c, N_{\mathbf{x}_{s'}}^2(x_{s, j})) + \sum_{j'=1}^{n_{s'}} \min(c, N_{\mathbf{x}_s}^2(x_{s', j'}))$$

where $N_{\mathbf{x}_{s'}}^1(x_{s, j}) = \sum_{j'=1}^{n_{s'}} \mathbf{1}[\|x_{s', j'} - x_{s, j}\| \leq R_1]$ and

$N_{\mathbf{x}_{s'}}^2(x_{s, j}) = \sum_{j'=1}^{n_{s'}} \mathbf{1}[R_1 < \|x_{s', j'} - x_{s, j}\| \leq R_2]$ are the number of between-subject neighbors over shorter and longer ranges, respectively.

We assign a bivariate lognormal distribution to (γ_1, γ_2) . Then

$$\begin{bmatrix} \log(\gamma_1) \\ \log(\gamma_2) \end{bmatrix} \sim \text{iid } N \left(\begin{bmatrix} \mu_{\gamma_1} \\ \mu_{\gamma_2} \end{bmatrix}, \begin{bmatrix} 1/\tau_{\gamma_1}^2 & \rho \\ \rho & 1/\tau_{\gamma_2}^2 \end{bmatrix} \right) \quad (8)$$

for precision parameters $\tau_{\gamma_1}^2$ and $\tau_{\gamma_2}^2$.

The hyperprior distributions are

$$\mu_{\gamma_1} \sim N(\mu_{10}, 1/\tau_{10}^2) \text{ and } \tau_{\gamma_1}^2 \sim \text{gamma}(\alpha_{10}, \beta_{10})$$

and

$$\mu_{\gamma_2} \sim N(\mu_{20}, 1/\tau_{20}^2) \text{ and } \tau_{\gamma_2}^2 \sim \text{gamma}(\alpha_{20}, \beta_{20})$$

where τ_{10}^2 and τ_{20}^2 are precision parameters and β_{10} and β_{20} are rate parameters.

4.3 Specifying the parameters and hyperparameters

We fix the values of the model parameters β , R_W , R_1 , R_2 , and c and of the hyperparameters ρ , μ_{10} , τ_{10}^2 , α_{10} , β_{10} , μ_{20} , τ_{20}^2 , α_{20} , and β_{20} .

We assume a constant intensity β across subjects. We apply our second stage model to four target brain regions (Figure 5) and found that the average intensity of points x does not vary greatly within each of these regions. We estimate β as the average number of activation points per subject per unit area in each of the brain regions, and we set β equal to 0.008 for Regions 1 and 2 and equal to 0.011 for Regions 3 and 4.

We set the within-subject hard-core distance R_W to be equal to 5mm, using the same reasoning as in the case of the first stage model. Based on findings by Frost and Goebel (2012) on misalignment in functional regions, we set $R_2 = 10$ mm. We set $R_1 = 5$ mm to be the half of R_2 . We set the saturation threshold c to be equal to 1.

The hyperparameter values are selected to ensure that the hyperpriors for the mean and precision parameters are uninformative. We set $\mu_{10} = 0.4$, $\tau_{10}^2 = 4$, $\mu_{20} = 0.8$, $\tau_{20}^2 = 2$. We set α_{10} and α_{20} equal to 0.5 and β_{10} and β_{20} equal to 0.1.

The correlation parameter ρ is updated iteratively during the initial iterations of the MCMC chain. For iteration l , ρ is estimated as the sampling correlation between the $\log(\gamma_{1,l})$ and $\log(\gamma_{1,l-1})$ of iteration $l-1$. The value of ρ is fixed once convergence is reached.

5 Application

We implement our first and second stage models via MCMC. We provide the details of our MCMC algorithms in Appendices A and B. Additional information on the implementation of the algorithms is provided in Appendix C.

5.1 Model fitting results: First stage model

As discussed in Section 2, data was collected on 21 individuals performing an attention network test. For our analysis, we selected an axial slice that intersects the frontal operculum cortex, the insular cortex, and the occipital cortex. This results in a set of 21 activation maps

for which z -statistics were calculated at each voxel location to reflect the strength of brain activation.

We fitted our first stage model to the maps of z -statistics to identify a configuration of activation units for each subject. The MCMC algorithm, described in Appendix A, was applied using the parameter and hyperparameter settings discussed in Section 3.3. We run an MCMC chain for a length of $L = 24000$ iterations for each subject and use a burn-in of 12000 iterations. Studying trace plots of the generated MCMC chains, we found that the height and size parameters mixed slowly. This is not surprising, given the high dimensionality of the posterior and the fact that the proposal density used to update the height and size parameters depends on the height and size of the point selected for updating. We display trace plots for the average height and size (over all the points for a given iteration) in Figure 2.

In Figure 3 we display the fitted activation surface along with the true surface for the same four individuals as in Figure 1. We see that there is a fair amount of subject to subject variability in the activation surfaces in terms of the spatial extent and amplitude of activation. Visual inspection suggests that the model does well picking up regions of activation.

Table 5.1 gives a summary of the configuration of activation units that was fitted for each subject. We calculate the summary measures of the fitted configurations at the subject level, *i.e.* the number of activation units, n_s , the average height $\bar{a}_s$, the average size $\bar{d}_s$, and we report the corresponding mean and standard deviations over all subjects. In addition, we calculate, for each subject, the maximum height $y_s^{\max} = \max(y_{s,v} | v \in V)$ of the activation surface and the average mean squared error $(\text{AMSE}) \frac{1}{n_s} \sum_{j=1}^{n_s} (y_{s,v} - A(v; \mathbf{x}))^2$, and we report the mean and standard deviation over all subjects. We see that the number of activation units fitted can vary greatly from subject to subject, allowing the first stage model to accurately reflect the between-subject variability in the spatial extent of activated areas as observed in Figure 3. We also note the high within-subject variability in the height and size parameters, indicating that the model is able to describe activation ranging from small regions with low amplitude to spatially extended activation regions with high amplitude. The average mean squared error is low relative to the average maximum height of activation, indicating that the model picks up activated regions well.

To check whether our model is consistent with the data, we simulate new datasets from our model. For subject s , we draw simulations from the posterior predictive distribution

$$p(y_s^{\text{rep}} | y_s) = \int p(y_s^{\text{rep}} | \mathbf{x}_s^M) p(\mathbf{x}_s^M | y_s) d\mathbf{x}_s^M,$$

and then compare the replicated dataset y_s^{rep} to the observed data y_s .

For model checking, we compare the observed data $y_{s,v}$ in each of the four primary areas of activation to the replicated data $y_{s,v}^{rep}$. The four regions are displayed in Figure 4. We calculate the average observed activation $\bar{y}_{s,v^*} = \sum_{v \in N_{v^*}} y_{s,v}$, where v^* is the central voxel for a neighborhood N_{v^*} of voxels. We define N_{v^*} to contain v^* and the 8 adjacent voxels to v^* . To assess the discrepancy between $\bar{y}_{s,v^*}^{rep}$ and $\bar{y}_{s,v^*}$, we calculate the probability that the test quantity replicated under the model exceeds the observed test quantity:

$$p_B = \iint \left(\sum_{s=1}^S I(\bar{y}_{s,v^*}^{rep} \geq \bar{y}_{s,v^*}) \right) \left(\prod_{s=1}^S p(y_s^{rep} | x_s^M) p(x_s^M | y_s) \right) dy^{rep} dx^M$$

where I is the indicator function. These probabilities are calculated as 0.51, 0.44, 0.51, and 0.51 for points 1 through 4 respectively, indicating that the sampled activation surface is representative of the observed activation surface.

5.2 Second stage model: Four regions of interest

For the second stage analysis, we used the configuration of activation locations x that maximized the first stage posterior distribution as input. Our goal with this analysis is to get insight into the inter-subject patterns of clustering that we observe in different functional areas of the brain. We identified four functional areas of interest which we label Regions 1 through 4. These regions are indicated on the axial slice in Figure 5. Regions 1 and 2 correspond to the left and right frontal operculum and insular cortices, respectively, and Regions 3 and 4 correspond to the left and right inferior lateral occipital cortex. The voxel of peak activation in Regions 1 through 4 has MNI (X, Y, Z)-coordinates $(-42, 19, -2)$, $(42, 6, -2)$, $(-52, -70, -2)$, and $(51, -71, -2)$, respectively.

We hypothesize that we will observe patterns of clustering in brain areas that engage in the experimental tasks. The occipital cortex (Regions 3 and 4) is involved in the visual processing of the target stimuli and the insular cortex (Regions 1 and 2) is involved in salience and network selection for the experimental tasks. Regions 1 and 2 are also part of the cingulo-opercular system of attention (Dosenbach *et al.*, 2008). This network is part of the ‘executive control’ system that is invoked by the fMRI task, and is hypothesized to be involved in maintaining vigilance of top-down attention across multiple trials. We would like to investigate how the degree and spatial scale of clustering vary among these regions.

5.3 Second stage model: Exploratory analysis

We start our second stage analysis by doing an exploratory analysis on the configuration of points x in each of the four regions. We are interested in studying the spatial scale at which we observe cross-subject clustering, and we calculate the pair correlation function for this purpose.

The pair correlation function is proportional to the first derivative of the Ripley’s K -function Ripley (1977). In particular, we calculate as the cross K -function, which is defined as

$$K_{ss'}(h) = \lambda_{s'}^{-1} E[\text{number of points of subject } s' \neq s \text{ within distance } h \text{ of a randomly chosen point of subject } s] \quad (9)$$

The corresponding pair correlation function is calculated based on taking the first derivative of $K_{ss'}$,

$$g_{ss'}(h) = \frac{1}{2\pi h} \frac{dK_{ss'}(h)}{dh}.$$

The cross g -function is calculated for each pair of subjects $(s_i, s_{i'}) \in S_m^*$. Then we calculate the pooled g -function

$$\bar{g}(h) = \sum_{(s, s') \in S_m^*} w_{ss'} g_{ss'}(h), \quad (10)$$

following Baddeley *et al.* (1993), for $w_{ss'} = n_s n_{s'} / \sum_{(t, t') \in S_m^*} n_t n_{t'}$.

For the pair correlation function, values greater than 1 indicate clustering. In Figure 6 we report the pooled $\bar{g}$ function calculated for each region. We calculate pointwise 95% envelopes under the assumption that the point patterns are independent across subjects. The envelopes are calculated by simulating independent realizations from the homogeneous Poisson process for the same number of subject pairs, m , as the data. Then the individual pair correlation functions $g_{ss'}$ and pooled function $\bar{g}$ are calculated for the m pairs.

Figure 6 shows patterns of inter-subject clustering over distance range up to 20mm. Note that the envelopes are quite narrow. This is due to averaging the pair correlation functions over the m pairs of subjects. We see that the range over which we observe clustering in Region 1 is longer than in the other regions. Regions 3 and 4 show shorter interaction ranges of about 15mm.

5.4 Second stage model: Model fitting results

We fitted our second stage model using the MCMC algorithm described in Appendix B. The parameter settings for this model was discussed and motivated in Section 4.

We fit our model to $K = 20$ sets of independent subject pairs and we run each MCMC chain for a length of $L = 5000$ iterations. Our chains show good mixing properties after about 200 iterations as shown in Figure 7.

We report the posterior mean of the inter-subject interaction $\exp(\mu_\gamma)$ for each of the four regions in Table 5.4. We calculate the exponent of μ_γ to interpret the parameter on the scale

of the γ_i 's, where $\gamma > 1$ indicates clustering and $\gamma < 1$ regularity. We report the posterior means of the precision parameters' reciprocal as well.

The summary in Table 5.4 indicates strong patterns of cross-subject clustering in each of the four regions. This result agrees with the exploratory analysis of the previous section. We note that cross-subject interactions are stronger for distance ranges from 5 to 10mm than for distance ranges of less or equal to 5mm, in each of the four regions. Also, the longer-range interactions are stronger in Regions 3 and 4 than in Regions 1 and 2.

5.5 Second stage model: Goodness-of-fit testing

To validate the fitted model, we compare the cross K -function (9) calculated for the data to 95% global rank envelopes (Myllymäki *et al.*, 2016) calculated based on the model. To calculate the global envelopes, we use the posterior predictive distribution

$$p(x_{\{1\}}^{rep}, \dots, x_{\{m\}}^{rep} | x_{\{1\}}, \dots, x_{\{m\}}) = \int \left(\prod_{i=1}^m p(x_{\{i\}}^{rep} | \gamma_{1i}, \gamma_{2i}) \right) \times p(\{\gamma_{1i}\}, \{\gamma_{2i}\}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2 | \{x_{\{i\}}\}) d\gamma_{11} \dots d\gamma_{1m} d\gamma_{21} \dots d\gamma_{2m} d\mu_{\gamma_1} d\tau_{\gamma_1}^2 d\mu_{\gamma_2} d\tau_{\gamma_2}^2$$

to generate realizations $x_{\{i\}}^{rep}$ from our fitted (null) model. This involves using posterior sampled $(\gamma_{1i}, \gamma_{2i})$ values and then generating point patterns corresponding to these values (using the algorithm given in Appendix B.4). We generate point patterns for m independent subject pairs. Then we calculate the cross K -function (9) for each subject pair and combine the cross K -functions into a pooled estimate, $\bar{K}$, using weights $w_{ss'} = n_s n_{s'}$ as in (10). We generate 1000 such pooled $\bar{K}$ -functions for a test at level $\alpha = 0.05$. We used the function `rank_envelope` from the R package `spptest` to calculate the global rank envelopes.

In Figure 8 we report variance-stabilized transformation of the data and envelope $\bar{K}$ -functions, $\bar{L}(h) = \{\bar{K}(h)/\pi\}^{1/2}$, for distance h . We see that the data function lines are within the envelopes for Regions 2, 3, and 4, indicating that the fitted model describes the data well. For Region 1, however, the data line above the upper envelope between distances of about 8 to 20mm.

The goodness-of-fit result suggests that the cut-off interaction distance of $R_2 = 10\text{mm}$ is too low for Region 1. We need to update our model to allow for interaction over longer distance ranges for Region 1.

6 Discussion

In this paper we developed a new approach for analyzing multi-subject fMRI data. Our approach is based on using a two stage model. For the first stage analysis, we proposed a marked spatial point process model to identify individual units of activation for each of the subjects in our study. For the second stage, we developed a pairwise interaction model to gain insight into how information is shared across subjects.

The two stage model has been developed to address shortcomings of fMRI multi-subject analysis methods and to illustrate new methods to summarize and characterize multi-subject brain activity. First, we address some shortcomings of standard fMRI methods by a) explicitly modeling the spatial features of the BOLD response (first stage model), b) accounting for the cross-subject misalignment in corresponding functional signals (second stage model), and c) by recognizing that information about activation should not be combined voxel-by-voxel across subjects.

Second, we introduce new methods of summarizing multi-subject data. In fMRI analyses, the focus is often on identifying the location of activation in order to map task activation to specific anatomical or functional landmarks. Here, our first stage model provides summaries of brain activation in terms of the magnitude, spatial extent and the number of activation units. Our second stage model studies inter-subject patterns in the spatial location of activation. The model is parameterized to give us insight into how likely it is that a subject has activation in a region of interest, given that other subjects in the group have activation in that region.

Our second stage analysis introduced a model with new features in spatial point process literature. The model is motivated by the Geyer saturation model. Eckel *et al.* (2009) expanded the Geyer saturation model for multitype processes. In this work we further develop the Geyer model by a) using inter-subject spatial neighborhoods as a means of accounting for cross-subject misalignment in activation, b) distinguishing between shorter and longer ranges of cross-subject interaction, and c) fitting our model to independent subject pairs. We recognized that the fitted interaction parameters vary by subject pair, and we used a Bayesian framework to model the variability in interaction parameters.

The second stage model is parameterized so that, given a subject has an activation point at location v , we get insight into how likely it is that another subject has an activation point in the neighborhood of v . The interaction parameter γ can be interpreted as follows: values of γ greater than one reflects that it is more likely for Subject s and Subject s' to have activation points in nearby locations than not. By estimating the interaction parameter we can compare how consistent activation is in different brain regions or across different groups of individuals. Lack of consistency in activation can be an indicator of aging or neural degeneration.

Studying the consistency in cross-subject activation ties in with work done by Rosenblatt *et al.* (2014) and Heller *et al.* (2007), who used methods to quantify the prevalence of activation across subjects at each location in the brain. Kannurpatti *et al.* (2010) were interested in how neural activity changes with aging. They found that there is an increase in the spatial variability of the activation location in older individuals compared to younger individuals, and more so under some experimental conditions than under others. While Kannurpatti *et al.* (2010) use somewhat ad hoc methods to study inter-subject variability, Xu *et al.* (2009) explicitly models the variability in subjects' activation units around a population activation unit.

We reported the results of fitting our two-stage model to data from an fMRI study in which subjects were asked to perform an attention network test. We found that our first stage model did well in describing the activation patterns of the 21 subjects in our study.

For the second stage analysis, we used as input the configuration of activation locations that maximized the first stage posterior. We identified four functional regions in the brain that are involved in the attention network task: the left and right frontal operculum and insular cortices, and the left and right inferior lateral occipital cortex. We hypothesized that we would observe patterns of inter-subject clustering in brain areas that engage in the experimental tasks. The model fitting results indicate that there is a strong degree of cross-subject clustering in each of the four regions, with the degree of clustering the highest in the left and right occipital cortices over a distance range of 5 to 10mm. The stronger clustering in the occipital cortex fits well with our scientific understanding that this area has a more specific function in performing tasks associated with the target conditions than the insula. The insula is more functionally heterogeneous than the regions of interest that we isolated in the occipital cortex (Nelson *et al.*, 2010; Kelly *et al.*, 2012). Based on previous findings (Madhyastha *et al.*, 2015b) it is expected that the right insula would be more consistently evoked than the left insula because of its presumed role in switching between networks (as is necessary when responding to the tasks of the attention network test). Over interaction ranges of 5 to 10mm we do indeed see stronger cross-subject interactions in the right insula. However, our goodness-of-fit results presented in Section 5.5 showed that our model is not a good fit for the left insula. We should be careful comparing the model results for left and right insula. Our exploratory analysis results did show, however, that we observe patterns of cross-subject clustering over a larger spatial scale in the left insula than in the right insula, which suggests more functional heterogeneity across subjects in the left insula for the task conditions of the study.

We have a couple of points of reflection on our choice of model for this analysis. First, we developed a model that uses as input the summary statistics derived at each voxel rather than the raw fMRI time-series data. A model could be developed based on the raw fMRI time-course data for each subject, but it would increase the computational cost of the first stage model substantially. It is standard in the literature to split up the analysis of the raw fMRI data into a first level (summarize the individual fMRI time-series at each voxel) and second level (identify activation foci) analysis. Also, for multi-subjects studies that compare the locations of activation across subjects (Xu *et al.*, 2009; Thirion *et al.*, 2006), it is typical to use a summary statistic at each voxel location as input, rather than the raw fMRI timecourse data.

Second, we used a two-stage model for our analysis. We did initially attempt to implement this analysis as a one-stage model, but the computational cost of running the MCMC algorithm was inhibitive given the computational resources available to us. We also found challenges in fine-tuning the one-stage model to discourage it from picking up too many activation centers, while reflecting the true strength of cross-subject interactions.

In future research, we would like to assess how sensitive second stage results are to our choice of first stage analysis method. In the interest of a unified modeling framework, we

used a point process model for our first stage analysis. However, various other methods can be considered to identify activation foci, for example Gaussian mixture modeling (Xu *et al.*, 2009), functional landmark modeling (Thirion *et al.*, 2010), and multidimensional scaling methods (Thirion *et al.*, 2006).

Even though our first stage model performed well in picking up the variability in the individual activation surfaces, we have identified some deficiencies. First, we made the simplifying assumption that the marks (height and size) are independent of the location parameters and that the marks are independent of each other. Studying correlations, using the fitted height and size parameters, suggests that prior knowledge of how the brain regions of interest is expected to engage with the task at hand is required to specify the model correlations structures accurately.

Second, we found that our first stage model is sensitive to our choice of lower and upper limits on the size parameter. If we specify both limits to have lower values, the model tends pick up a larger number of more peaked activation points, and, conversely, if we specify the limits to have higher values, the model picks up fewer and larger activation units. We found that our choice of size parameter limits is in line with other studies the literature that also aim to indentify foci or centers of activation (Thirion *et al.*, 2010; Keller *et al.*, 2009; Xu *et al.*, 2009; Kannurpatti *et al.*, 2010). A fuller assessment is warranted, but it is not within the scope of the current work.

Third, we found that the high dimensionality of first stage posterior distribution makes the MCMC algorithm slow to converge. In Appendix C we give instructions to help speed up convergence.

Along these lines, as mentioned in Appendix C, the implementation of the second stage model MCMC algorithm can be slowed down substantially by the auxiliary variable method. In future work, to speed up computation, we would consider implementing the second stage model using pseudolikelihood estimation with estimates obtained using existing software for generalised linear mixed models (Illian and Hendrichsen, 2010).

In summary, the model that we introduce in this paper allows us to study how information about activation is shared in different functional areas in the brain. We highlighted how our model can be used to summarize and characterize multi-subject fMRI data for an attention network test. A promising route for future analyses would be to use these methods to characterize brain activations elicited by different categories of experimental conditions. We also envision the further development of our model to introduce methods that assign group membership to individuals based on the characteristic interaction patterns of each group.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research was supported by grants from the National Institutes of Health (1RC4NS073008-01 and P50NS062684), by the National Institute of Environmental Health Sciences of the National Institutes of Health

(T32ES015459), and by the Knut and Alice Wallenberg Foundation (KAW 2012.0067) and the Swedish Foundation of Strategic Research (SSF AM13-0066).

Appendices

Here we describe the algorithms that we will use to sample from the posterior distributions of the first and second stage models.

A MCMC algorithm for first stage model

To sample from $p(\mathbf{x}^M | \mathbf{y})$, we developed a Metropolis-Hastings algorithm based on the algorithm proposed in Geyer and Moller (1994) and Moller and Waagepetersen (2003). We update the configuration $\mathbf{x}^M$ at each iteration of the algorithm until convergence is reached. To update $\mathbf{x}^M$, we propose the addition, deletion or moving of a point.

Algorithm

For each of $s = 1, \dots, S$ iterate through the following steps until convergence is reached:

1. sample $r \sim \text{unif}[0, 1]$
2. if $r < p_b$: propose the addition of a point to $\mathbf{x}_s^M$
 if $p_b < r < (p_b + p_d)$: propose the deletion of a point of $\mathbf{x}_s^M$
 else: propose moving a point in $\mathbf{x}_s^M$ with probability $p_m = 1 - (p_b + p_d)$

For each of the birth, death, and move steps we use proposal densities q_b , q_d and q_m , respectively. Let $\mathbf{x}^{M'}$ denote a new configuration that is based on the current configuration $\mathbf{x}^M$. Then, the probability with which we accept each of the steps is, respectively,

$$\alpha_b(\mathbf{x}^M, \mathbf{x}^{M'}) = \min \left(\frac{p(\mathbf{x}^{M'} | \mathbf{y}) q_d(\mathbf{x}^M | \mathbf{x}^{M'}) p_d}{p(\mathbf{x}^M | \mathbf{y}) q_b(\mathbf{x}^{M'} | \mathbf{x}^M) p_b}, 1 \right)$$

$$\alpha_d(\mathbf{x}^M, \mathbf{x}^{M'}) = \min \left(\frac{p(\mathbf{x}^{M'} | \mathbf{y}) q_b(\mathbf{x}^M | \mathbf{x}^{M'}) p_b}{p(\mathbf{x}^M | \mathbf{y}) q_d(\mathbf{x}^{M'} | \mathbf{x}^M) p_d}, 1 \right)$$

$$\alpha_m(\mathbf{x}^M, \mathbf{x}^{M'}) = \min \left(\frac{p(\mathbf{x}^{M'} | \mathbf{y}) q_m(\mathbf{x}^M | \mathbf{x}^{M'})}{p(\mathbf{x}^M | \mathbf{y}) q_m(\mathbf{x}^{M'} | \mathbf{x}^M)}, 1 \right)$$

Insertion of a point

To sample a new point $\xi = (x_\xi, a_\xi, d_\xi)$, we choose a proposal density function q_b that is proportional to the Papangelou conditional intensity $p(\mathbf{x}^M \cup \xi | \mathbf{y}) / p(\mathbf{x}^M | \mathbf{y})$.

We propose the new location, height, and size parameters sequentially,

$$q_b(x^M \cup \xi | x^M) = q_b(x_\xi | a_0, d_0, x^M, y) q_b(a_\xi | x_\xi, d_0, x^M, y) q_b(d_\xi | x_\xi, a_\xi, x^M, y).$$

In order to propose new parameter values in turn, we use the fixed values a_0 and d_0 . We set a_0 and d_0 equal to typical values of a and d , and let $a_0 = 1.3$ and $d_0 = 22\text{mm}^2$.

The birth proposal for location x_ξ is set to be proportional to the Papangelou conditional intensity,

$$q_b(x_\xi | a_0, d_0, x^M, y) \propto \frac{p(x \cup x_\xi) p(y | x^M \cup (x_\xi, a_0, d_0))}{p(x) p(y | x^M)}.$$

We assume that the new point ξ is added to the point configuration of subject 1. Then

$$\frac{p(x \cup x_\xi)}{p(x)} = \beta(x_\xi) \text{ if } \|x_{1j} - x_\xi\| > R_W \forall j \text{ and zero otherwise.}$$

The updated activation surface of subject 1 is modeled as

$$A_{1,v}^\xi = h(v; (x_\xi, a_0, d_0)) + h(v; x_{s,1}^M) + \dots + h(v; x_{s,n_s}^M) = h(v; (x_\xi, a_0, d_0)) + A_{1,v}$$

and the likelihood given the updated point configuration is

$$\begin{aligned} p(y | x^M \cup (x_\xi, a_0, d_0)) &\propto \exp\left(-\frac{1}{2\tau^2} \left(\sum_{v \in V} (y_{1,v} - A_{1,v}^\xi)^2 + \sum_{s=2}^S \sum_{v \in V} (y_{s,v} - A_{s,v})^2 \right)\right) \\ &= \exp\left(-\frac{1}{2\tau^2} \left(\sum_{v \in V} (y_{1,v} - A_{1,v} - h(v; (x_\xi, a_0, d_0)))^2 \right)\right) \times \exp\left(-\frac{1}{2\tau^2} \sum_{s=2}^S \sum_{v \in V} (y_{s,v} - A_{s,v})^2\right). \end{aligned}$$

Then the ratio of likelihoods simplifies to

$$\begin{aligned}
\frac{p(\mathbf{y}|\mathbf{x}^M \cup (x_\xi, a_0, d_0))}{p(\mathbf{y}|\mathbf{x}^M)} &= \exp\left(-\frac{1}{2\tau^2} \sum_v ((y_{1,v} - A_{1,v} - h(v; (x_\xi, a_0, d_0)))^2 - (y_{1,v} - A_{1,v})^2)\right) \\
&= \exp\left(-\frac{1}{2\tau^2} \sum_v (h(v; (x_\xi, a_0, d_0))^2 - 2h(v; (x_\xi, a_0, d_0))(y_{1,v} - A_{1,v}))\right) \\
&\simeq \exp\left(-\frac{1}{2\tau^2} \left(\frac{a_0^2 d_0}{2\Delta_{v1}\Delta_{v2}\log 2} - 2 \sum_v h(v; (x_\xi, a_0, d_0))(y_{1,v} - A_{1,v})\right)\right) \\
&\propto \exp\left(-\frac{1}{\tau^2} \left(\sum_v h(v; (x_\xi, a_0, d_0))(y_{1,v} - A_{1,v})\right)\right).
\end{aligned}
\tag{11}$$

The above simplification is derived as follows

$$\sum_v h(v; (x, a, d))^2 \Delta_{v1} \Delta_{v2} \simeq \int h(v; (x, a, d))^2 dv_1 dv_2 = a^2 d / (2 \log 2)$$

Here v_1 and v_2 refer to the voxel sides and Δ_{v1} and Δ_{v2} are the length of the voxel sides in millimeters.

To sample from (11) we need to evaluate the conditional density function at each location on the grid and use rejection sampling. The Fast Fourier Transform algorithm can be used to speed up the evaluation of (11).

To propose a height of the new point, we let

$$\begin{aligned}
q_b(a_\xi | x_\xi, d_0, \mathbf{x}^M, \mathbf{y}) &\propto \frac{p(a_\xi) p(\mathbf{y}|\mathbf{x}^M \cup (x_\xi, a_\xi, d_0))}{p(\mathbf{y}|\mathbf{x}^M)} \\
&\propto p(a_\xi) \exp\left(-\frac{1}{2\tau^2} \left(a_\xi^2 \frac{d_0}{2\Delta_{v1}\Delta_{v2}\log 2} - 2a_\xi \sum_v h(v; x_\xi, 1, d_0)(y_{1,v} - A_{1,v})\right)\right) \\
&\propto p(a_\xi) p_N(a_\xi; x_\xi, d_0, \mathbf{x}^M, \mathbf{y})
\end{aligned}
\tag{12}$$

Here p_N is the density of a normal distribution with mean

$$\frac{2\Delta_{v1}\Delta_{v2}\log 2}{d_0} \sum_v h(v; x_\xi, 1, d_0)(y_{1,v} - A_{1,v}) \text{ and variance } \frac{2\tau^2 \Delta_{v1}\Delta_{v2}\log 2}{d_0}.$$

We use rejection sampling to sample from (12).

We propose a value for the new size parameter based on the ratio

$$\frac{p(\mathbf{y}|\mathbf{x}^M \cup (x_\xi, a_\xi, d_\xi))}{p(\mathbf{y}|\mathbf{x}^M)} \propto \exp\left[-\frac{1}{2\tau^2}\left(\frac{a_\xi^2 d_\xi}{2\Delta_{v1}\Delta_{v2}\log 2} - 2\sum_v h(v; (x_\xi, a_\xi, d_\xi))(y_{1,v} - A_{1,v})\right)\right] \quad (13)$$

for $d_\xi \in [d_{low}, d_{up}]$. This ratio is used to derive a piecewise log-linear proposal function

$$q_b(d_\xi | x_\xi, a_\xi, \mathbf{x}^M, \mathbf{y}) \propto \exp[p_{i-1} + (p_i - p_{i-1}) \frac{d_\xi - \delta_{i-1}}{\delta_i - \delta_{i-1}}]$$

for $d_{low} = \delta_0 < \delta_1 < \dots < \delta_m = d_{up}$ and p_i the logarithm of ratio (13), evaluated at δ_i ,

$$p_i = \log \left(\frac{p(\mathbf{y}|\mathbf{x}^M \cup (x_\xi, a_\xi, \delta_i))}{p(\mathbf{y}|\mathbf{x}^M)} \right).$$

Removal of a point

We propose to remove point $x_{s,j}^M$ from the point configuration of subject s with probability

$$q_d(\mathbf{x}_{-(s,j)}^M | \mathbf{x}^M) = 1/n_s, \text{ for } \mathbf{x}_{-(s,j)}^M = \mathbf{x}^M \setminus x_{s,j}^M$$

Moving a point

To update an existing point, we randomly choose one of the n_s points of subject s . To simplify notation, let's assume that we selected the first point of subject 1,

$x_{1,1}^M = (x_{1,1}, a_{1,1}, d_{1,1})$. Then, to obtain a good acceptance rate, we set our proposal density to

be proportional to the Papangelou conditional intensity $p(\mathbf{x}_{-(1,1)}^M \cup x_{1,1}^M | \mathbf{y}) / p(\mathbf{x}^M | \mathbf{y})$. Here

$\mathbf{x}_{-(1,1)}^M$ denotes the full marked point configuration $\mathbf{x}^M$ excluding point $x_{1,1}^M$.

Similar to the birth proposal, we update the location, height and size parameters sequentially,

$$q_m(\mathbf{x}_{-(s,j)}^M \cup (x_\xi, a_\xi, d_\xi) | \mathbf{x}^M) = q_m(x_\xi | a_0, d_0, \mathbf{x}^M, \mathbf{y}) q_m(a_\xi | x_\xi, d_0, \mathbf{x}^M, \mathbf{y}) q_m(d_\xi | x_\xi, a_\xi, \mathbf{x}^M, \mathbf{y})$$

We specify the proposal density for location x_ξ to be proportional to the Papangelou conditional intensity

$$q_m(x_\xi | a_0, d_0, \mathbf{x}^M, \mathbf{y}) \propto \frac{p(\mathbf{x}_{-(1,1)} \cup x_\xi) p(\mathbf{y} | \mathbf{x}_{-(1,1)}^M \cup (x_\xi, a_0, d_0))}{p(\mathbf{x}) p(\mathbf{y} | \mathbf{x}^M)} \quad (14)$$

Using the same steps as in the case of the birth proposal, we derive the proposal density

$$q_m(x_\xi | a_0, d_0, \mathbf{x}^M, \mathbf{y}) \propto \frac{\beta(x_\xi)}{\beta(x_{1,1})} \exp\left(\frac{1}{\tau^2} \left(\sum_v h(v; x_\xi, a_0, d_0)(y_{1,v} - A_{1,v}^{-(1,1)}) \right)\right) \quad (15)$$

if $\|x_{1j} - x_\xi\| > R_W$ for $j = 2, \dots, n_1$ and zero otherwise. Here $A_{1,v}^{-(1,1)} = A_{1,v} - h(v; x_{1,1}^M)$ is the fitted activation surface for subject 1 after removing point $x_{1,1}^M$. As before, we use fixed values a_0 and d_0 to allow us to propose a move in location, size and height in turn. We set $a_0 = 1.3$ and $d = 22\text{mm}^2$.

To propose a height at the new location x_ξ , we let

$$q_m(a_\xi | x_\xi, d_0, \mathbf{x}^M, \mathbf{y}) \propto \frac{p(a_\xi)p(\mathbf{y} | \mathbf{x}_{-(1,1)}^M \cup (x_\xi, a_\xi, d_0))}{p(a_{1,1})p(\mathbf{y} | \mathbf{x}^M)} \propto \frac{p(a_\xi)}{p(a_{1,1})} p_N(a_\xi; x_\xi, d_0, \mathbf{x}_1^M | \mathbf{x}_{1,1}^M, \mathbf{y}) \quad (16)$$

Here p_N is the density of a normal distribution with mean

$$\frac{2\Delta_{v1}\Delta_{v2}\log 2}{d_0} \sum h(v; x_\xi, 1, d_0)(y_{1,v} - A_{1,v}^{-(1,1)}) \text{ and variance } \frac{2\tau^2\Delta_{v1}\Delta_{v2}\log 2}{d_0}.$$

To propose a new value for the height parameter, we follow the same steps as for the birth update, except that p_i is now calculated as

$$p_i = \log\left(\frac{p(\mathbf{y} | \mathbf{x}_{-(1,1)}^M \cup (x_\xi, a_\xi, \delta_i))}{p(\mathbf{y} | \mathbf{x}^M)}\right) = -\frac{1}{2\tau^2} \left(\frac{a_\xi^2 \delta_i}{2\Delta_{v1}\Delta_{v2}\log 2} - 2 \sum_v h(v; x_\xi, a_\xi, \delta_i)(y_{1,v} - A_{1,v}^{-(1,1)}) \right)$$

B MCMC algorithm for the second stage model

In the second stage model, we sample from the posterior

$p(\{\gamma_{1i}\}, \{\gamma_{2i}\}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2 | \{\mathbf{x}_{\{i\}}\})$. To simplify notation, we do not indicate dependence on the model parameters and hyperparameters that we fix (discussed in Section 4.3). We use the notation $\{\gamma_{1i}\}$, $\{\gamma_{2i}\}$, and $\{\mathbf{x}_{\{i\}}\}$ to indicate the full set of m variables.

We iterate through the following steps until convergence is reached:

1. Sample from $p(\gamma_{1i}, \gamma_{2i} | \mathbf{x}_{\{i\}}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2)$, for $i = 1, \dots, m$.
2. Sample from $p(\mu_{\gamma_1} | \{\gamma_{1i}\}, \{\mathbf{x}_{\{i\}}\}, \tau_{\gamma_1}^2)$

3. Sample from $p(\tau_{\gamma_1}^2 | \{\gamma_{1i}\}, \{\mathbf{x}_{\{i\}}\}, \mu_{\gamma_1})$
4. Sample from $p(\mu_{\gamma_2} | \{\gamma_{2i}\}, \{\mathbf{x}_{\{i\}}\}, \tau_{\gamma_2}^2)$
5. Sample from $p(\tau_{\gamma_2}^2 | \{\gamma_{2i}\}, \{\mathbf{x}_{\{i\}}\}, \mu_{\gamma_2})$

Below we discuss the details of each of the above steps.

B.1 Sampling from the posterior distribution of (γ_1, γ_2)

We sample from the joint distribution of (γ_1, γ_2) . Dropping the subscript i to simplify notation, we accept a proposed update γ'_1, γ'_2 with probability

$$\alpha(\gamma_1, \gamma_2; \gamma'_1, \gamma'_2) = \min \left(\frac{p(\gamma'_1, \gamma'_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma_1, \gamma_2 | \gamma'_1, \gamma'_2)}{p(\gamma_1, \gamma_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma'_1, \gamma'_2 | \gamma_1, \gamma_2)}, 1 \right)$$

The posterior of (γ_1, γ_2) is proportional to $p(\mathbf{x} | \gamma_1, \gamma_2) p(\gamma_1, \gamma_2 | \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2)$, the densities given in (7) and (8). We write $p(\mathbf{x} | \gamma_1, \gamma_2) = f_{\gamma_1, \gamma_2}(\mathbf{x}) / Z_{\gamma_1, \gamma_2}$ to emphasize that the density of $\mathbf{x}$ depends on an intractable normalizing constant Z_{γ_1, γ_2} .

To avoid direct calculation of the normalizing constants, we use the auxiliary variable method proposed by Berthelsen and Møller (2006). For this method, the authors introduced an auxiliary variable $\mathbf{v}$ defined on the same state space as $\mathbf{x}$. The corresponding acceptance probability is

$$\alpha(\gamma_1, \gamma_2, \mathbf{v}; \gamma'_1, \gamma'_2, \mathbf{v}') = \min \left(\frac{p(\gamma'_1, \gamma'_2, \mathbf{v}' | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma_1, \gamma_2, \mathbf{v} | \gamma'_1, \gamma'_2, \mathbf{v}')}{p(\gamma_1, \gamma_2, \mathbf{v} | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma'_1, \gamma'_2, \mathbf{v}' | \gamma_1, \gamma_2, \mathbf{v})}, 1 \right)$$

The posterior of $(\gamma_1, \gamma_2, \mathbf{v})$ is proportional to

$$p(\mathbf{v} | \mathbf{x}, \gamma_1, \gamma_2, \mu_{\gamma_1}, \tau_{\gamma_1}^2) p(\mathbf{x} | \gamma_1, \gamma_2) p(\gamma_1, \gamma_2 | \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2).$$

The auxiliary density of $\mathbf{v}$ has to be chosen to ensure good mixing properties for the algorithm. Berthelsen and Møller (2006) suggest the density of $\mathbf{v}$ to be the same as the density of $\mathbf{x}$, with the parameters γ_1 and γ_2 fixed to values $\tilde{\gamma}_1$ and $\tilde{\gamma}_2$ that the posterior is concentrated around. As a result we let the density of $\mathbf{v}$ be,

$$p(\mathbf{v} | \mathbf{x}, \gamma_1, \gamma_2, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) = p(\mathbf{v} | \tilde{\gamma}_1, \tilde{\gamma}_2) = f_{\tilde{\gamma}_1, \tilde{\gamma}_2}(\mathbf{v}) / Z_{\tilde{\gamma}_1, \tilde{\gamma}_2}.$$

The proposal density for $(\mathbf{v}, \gamma_1, \gamma_2)$ is simplified to

$$q(\mathbf{v}', \gamma'_1, \gamma'_2 | \mathbf{v}, \gamma_1, \gamma_2) = q(\gamma'_1, \gamma'_2 | \mathbf{v}, \gamma_1, \gamma_2) q(\mathbf{v}' | \gamma'_1, \gamma'_2, \mathbf{v}, \gamma_1, \gamma_2) = q(\gamma'_1, \gamma'_2) q(\mathbf{v}' | \gamma'_1, \gamma'_2)$$

Following Berthelsen and Møller (2006), we define $q(\mathbf{v} | \gamma_1, \gamma_2) = f_{\gamma_1, \gamma_2}(\mathbf{v}) / Z_{\gamma_1, \gamma_2}$ so that the normalizing constants cancel out in the Hastings ratio. The proposal density for (γ_1, γ_2) is discussed in Appendix C. Incorporating the above constructions, the Hastings ratio simplifies to

$$\begin{aligned} & \frac{p(\gamma'_1, \gamma'_2, \mathbf{v}' | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma_1, \gamma_2, \mathbf{v} | \gamma'_1, \gamma'_2, \mathbf{v}')}{p(\gamma_1, \gamma_2, \mathbf{v} | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma'_1, \gamma'_2, \mathbf{v}' | \gamma_1, \gamma_2, \mathbf{v})} \\ &= \frac{p(\mathbf{x} | \gamma'_1, \gamma'_2), p(\mathbf{v}' | \bar{\gamma}_1, \bar{\gamma}_2) p(\gamma'_1, \gamma'_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma_1, \gamma_2) q(\mathbf{v} | \gamma_1, \gamma_2)}{p(\mathbf{x} | \gamma_1, \gamma_2) p(\mathbf{v} | \bar{\gamma}_1, \bar{\gamma}_2) p(\gamma_1, \gamma_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma'_1, \gamma'_2) q(\mathbf{v}' | \gamma'_1, \gamma'_2)} \\ &= \frac{f_{\gamma'_1, \gamma'_2}(\mathbf{x}) Z_{\gamma'_1, \gamma'_2} f_{\bar{\gamma}_1, \bar{\gamma}_2}(\mathbf{v}') Z_{\bar{\gamma}_1, \bar{\gamma}_2} p(\gamma'_1, \gamma'_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma_1, \gamma_2) f_{\gamma_1, \gamma_2}(\mathbf{v}) Z_{\gamma_1, \gamma_2}}{f_{\gamma_1, \gamma_2}(\mathbf{x}) Z_{\gamma_1, \gamma_2} f_{\bar{\gamma}_1, \bar{\gamma}_2}(\mathbf{v}) Z_{\bar{\gamma}_1, \bar{\gamma}_2} p(\gamma_1, \gamma_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma'_1, \gamma'_2) f_{\gamma'_1, \gamma'_2}(\mathbf{v}') Z_{\gamma'_1, \gamma'_2}} \\ &= \frac{f_{\gamma'_1, \gamma'_2}(\mathbf{x}) f_{\bar{\gamma}_1, \bar{\gamma}_2}(\mathbf{v}') p(\gamma'_1, \gamma'_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma_1, \gamma_2) f_{\gamma_1, \gamma_2}(\mathbf{v})}{f_{\gamma_1, \gamma_2}(\mathbf{x}) f_{\bar{\gamma}_1, \bar{\gamma}_2}(\mathbf{v}) p(\gamma_1, \gamma_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\gamma'_1, \gamma'_2) f_{\gamma'_1, \gamma'_2}(\mathbf{v}')} \end{aligned}$$

B.2 Sampling from the posterior distribution of μ_{γ_1} and μ_{γ_2}

Here we derive the posterior distribution for μ_{γ_1} only. The posterior distribution of μ_{γ_2} is derived the same way.

Assuming a bivariate lognormal distribution for (γ_1, γ_2) and a normal distribution for μ_{γ_1} , as discussed in Section 4.2, we derive the posterior as follows

$$\begin{aligned} & p(\mu_{\gamma_1} | \{\gamma_{1i}\}, \{\gamma_{2i}\}, \mu_{\gamma_2}, \tau_{\gamma_1}^2, \tau_{\gamma_2}^2) \\ & \propto p(\mu_{\gamma_1} | \mu_{10}, \tau_{10}^2) \prod_{i=1}^m p(\gamma_{1i}, \gamma_{2i} | \mu_{\gamma_1}, \mu_{\gamma_2}, \tau_{\gamma_1}^2, \tau_{\gamma_2}^2) \\ & \propto \exp\left(-\frac{\tau_{10}^2}{2}(\mu_{\gamma_1} - \mu_{10})^2\right) \exp\left(-\frac{\tau_{\gamma_1}^2}{2(1-\rho^2)} \sum_{i=1}^m (\ln \gamma_{1i} - \mu_{\gamma_1})^2 + \frac{\tau_{\gamma_1} \tau_{\gamma_2} \rho}{1-\rho^2} \sum_{i=1}^m (\ln \gamma_{1i} - \mu_{\gamma_1})(\ln \gamma_{2i} - \mu_{\gamma_2})\right) \\ & \propto \exp\left(-\frac{1}{2} \mu_{\gamma_1}^2 \left(\tau_{10}^2 + \frac{m \tau_{\gamma_1}^2}{1-\rho^2}\right) - \frac{1}{2} \mu_{\gamma_1} \left(-2 \tau_{10}^2 \mu_{10} - 2 \frac{\tau_{\gamma_1}^2}{1-\rho^2} \sum_{i=1}^m \ln \gamma_{1i} - 2 \frac{\tau_{\gamma_1} \tau_{\gamma_2} \rho}{1-\rho^2} (m \mu_{\gamma_2} - \sum_{i=1}^m \ln \gamma_{2i})\right)\right) \end{aligned}$$

From the above, it follows that

$$\mu_{\gamma_1} | \{\gamma_{1i}\}, \{\gamma_{2i}\}, \mu_{\gamma_2}, \tau_{\gamma_1}^2, \tau_{\gamma_2}^2 \sim N\left(\tau_{10}^2 \mu_{10} + \frac{\tau_{\gamma_1}^2}{1-\rho^2} \sum_{i=1}^m \ln \gamma_{1i} + \frac{\tau_{\gamma_1} \tau_{\gamma_2} \rho}{1-\rho} (m \mu_{\gamma_2} - \sum_{i=1}^m \ln \gamma_{2i}) / (\tau_{10}^2 + \frac{m \tau_{\gamma_1}^2}{1-\rho^2}), 1 / (\tau_{10}^2 + \frac{m \tau_{\gamma_1}^2}{1-\rho^2})\right)$$

B.3 Sampling from the posterior distribution of $\tau_{\gamma_1}^2$ and $\tau_{\gamma_2}^2$

We use a Metropolis-Hastings step to sample from the posterior of $\tau_{\gamma_1}^2$. We accept the proposed $\tau_{\gamma_1}^{2'}$ with probability

$$\alpha(\tau_{\gamma_1}^2; \tau_{\gamma_1}^{2'}) = \min \left(\frac{p(\tau_{\gamma_1}^{2'} | \{\gamma_{1i}\}, \{\gamma_{2i}\}, \mu_{\gamma_1}, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\tau_{\gamma_1}^2 | \tau_{\gamma_1}^{2'})}{p(\tau_{\gamma_1}^2 | \{\gamma_{1i}\}, \{\gamma_{2i}\}, \mu_{\gamma_1}, \mu_{\gamma_2}, \tau_{\gamma_2}^2) q(\tau_{\gamma_1}^{2'} | \tau_{\gamma_1}^2)}, 1 \right).$$

Here the posterior of $\tau_{\gamma_1}^2$ is proportional to

$$p(\tau_{\gamma_1}^2 | \alpha_0, \beta_0) \prod_{i=1}^m p(\gamma_{1i}, \gamma_{2i} | \mu_{\gamma_1}, \mu_{\gamma_2}, \tau_{\gamma_1}^2, \tau_{\gamma_2}^2).$$

We derive a proposal density that ensures good mixing properties for the algorithm.

Assuming that $\gamma_{1i} \sim \text{logn}(\mu_{\gamma_1}, \tau_{\gamma_1}^2)$ (and that γ_{1i} and γ_{2i} are independent) and

$\tau_{\gamma_1}^2 \sim \text{gamma}(\alpha_0, \beta_0)$, the posterior for $\tau_{\gamma_1}^2$ would be $\text{gamma}(\alpha_0 + \frac{m}{2}, \beta_0 + \frac{1}{2} \sum_{i=1}^m (\ln \gamma_{1i} - \mu_{\gamma_1})^2)$.

We use this density as the proposal density $q(\tau_{\gamma_1}^2)$.

B.4 Generating samples from the proposal density $q(v/\gamma_1, \gamma_2)$

To implement the auxiliary variable method, we need to generate samples from model (7).

We generate samples of v by proposing the insertion or deletion of a point with probabilities p_b and p_d respectively. Then, the probability with which we accept the insertion or deletion of a point is, respectively,

$$\alpha_b(v, v') = \min \left(\frac{p(v' | \gamma_1, \gamma_2) q_d(v | v') p_d}{p(v | \gamma_1, \gamma_2) q_b(v' | v) p_b}, 1 \right)$$

$$\alpha_d(\mathbf{v}, \mathbf{v}') = \min \left(\frac{p(\mathbf{v}'|\gamma_1, \gamma_2)q_b(\mathbf{v}|\mathbf{v}')p_b}{p(\mathbf{v}|\gamma_1, \gamma_2)q_d(\mathbf{v}'|\mathbf{v})p_d}, 1 \right). \quad (17)$$

Insertion of a point

To add a point v_ξ to the existing configuration, we use a proposal density q_b that is proportional to the Papangelou conditional intensity. Let us assume that subjects s and s' have configurations $\mathbf{v}_s = \{v_{s,1}, \dots, v_{s,n_s}\}$ and $\mathbf{v}_{s'} = \{v_{s',1}, \dots, v_{s',n_{s'}}\}$, respectively, that $\mathbf{v} = \{\mathbf{v}_s, \mathbf{v}_{s'}\}$, and that the point v_ξ is added to the point configuration of subject s . Then for $\mathbf{v}' = \mathbf{v} \cup v_\xi$

$$q_b(\mathbf{v}'|\mathbf{v}) = \frac{p(\mathbf{v} \cup v_\xi)}{p(\mathbf{v})} = \beta \gamma_1^{s_1(\mathbf{v} \cup v_\xi) - s_1(\mathbf{v})} \gamma_2^{s_2(\mathbf{v} \cup v_\xi) - s_2(\mathbf{v})}$$

for $\|\mathbf{v}_{s,j} - v_\xi\| > R_W$ for $j = 1, 2, \dots, n_s$ and zero otherwise.

Then,

$$\begin{aligned} & s_1(\mathbf{v} \cup v_\xi) - s_1(\mathbf{v}) \\ &= \sum_{j=1}^{n_1} \min(c, N_{\mathbf{v}_{s'}}^1(v_{s,j})) + \min(c, N_{\mathbf{v}_{s'}}^1(v_\xi)) + \sum_{j'=1}^{n_{s'}} \min(c, N_{\mathbf{v}_s \cup v_\xi}^1(v_{s',j'})) \\ & \quad - \sum_{j=1}^{n_1} \min(c, N_{\mathbf{v}_{s'}}^1(v_{s,j})) - \sum_{j'=1}^{n_{s'}} \min(c, N_{\mathbf{v}_s}^1(v_{s',j'})) \\ &= \min(c, N_{\mathbf{v}_{s'}}^1(v_\xi)) + \sum_{j'=1}^{n_{s'}} \delta_{\mathbf{v}_s \cup v_\xi}^1(v_{s',j'}). \end{aligned}$$

where

$$\delta_{\mathbf{v}_s \cup v_\xi}^1(v_{s',j'}) = \begin{cases} 1 & \text{if } N_{\mathbf{v}_s}^1(v_{s',j'}) < N_{\mathbf{v}_s \cup v_\xi}^1(v_{s',j'}) \leq d, \\ 0 & \text{otherwise.} \end{cases}$$

Here $N_{\mathbf{v}_s}^1(v_{s',j'})$ is the number of short-range between-subject neighbors that the point $v_{s',j'}$ has in the set of points $\mathbf{v}_s$.

Similarly,

$$s_2(\mathbf{v} \cup v_\xi) - s_2(\mathbf{v}) = \min(c, N_{\mathbf{v}_s}^2(v_\xi)) + \sum_{j'=1}^{n_{s'}} \delta_{\mathbf{v}_s \cup v_\xi}^2(v_{s',j'})$$

where

$$\delta_{\mathbf{v}_s \cup \mathbf{v}_{\xi}}^2(v_{s'}, j') = \begin{cases} 1 & \text{if } N_{\mathbf{v}_s}^2(v_{s'}, j') < N_{\mathbf{v}_s \cup \mathbf{v}_{\xi}}^2(v_{s'}, j') \leq d, \\ 0 & \text{otherwise,} \end{cases}$$

and $N_{\mathbf{v}_s}^2(v_{s'}, j')$ is the number of longer-range between-subject neighbors.

Deletion of a point

We propose to remove point $v_{s,j}$ from the point configuration of subject s with probability $q_d(\mathbf{v} \setminus v_{s,j} | \mathbf{v}) = 1/n_s$.

C Implementation

The MCMC algorithms for the first and second stage models were both implemented in R (R Core Team, 2015). The package `smoothie` was used to implement the Fast Fourier Transform, packages `spatstat` and `fields` were used to calculate second order summary statistics and generate graphs, and `spptest` to implement the global envelope test.

For the first stage model, we found that generating a good starting point for our algorithm can greatly reduce the number of iteration required to reach convergence. Initially, we generated a starting configuration of 2 activation units per subject (locations chosen randomly in V), but it took a very large number of iterations to reach convergence from this starting point.

We used the following two steps to create starting configuration for each subject. First, we select two point randomly and set $p_b = 0.9$ and $p_d = 0.1$ to encourage new points to be added to the point configuration of each subject. We use these settings for the first 1,500 iterations of the algorithm.

Second, for iterations 1,501 through 7,500, we set $p_b = 0.167$ and $p_d = 0.167$. During these iterations, the initial points added to the configuration (that are typically large and high), are being updated to be more representative of the activation surface. This means, on average, that the height and size of existing points are updated to become smaller and that more points are added to the configuration. Note that in the above two steps, we used $a_0 = 1.3$ and $d_0 = 22$ for the insertion, removal, and updating of points.

We use the configuration generated during the final iteration above as a starting point for our algorithm. For our main run of the algorithm, we use the settings $p_b = 0.167$ and $p_d = 0.167$. For the insertion and removal of points, we set $a_0 = 1.3$ and $d_0 = 22$ and for updating points, we set $a_0 = a_{1,1}$ and $d_0 = d_{1,1}$ where $a_{1,1}$ and $d_{1,1}$ are the height and size, respectively, of the point that has been selected to be updated. We run the algorithm for 24,000 iterations. We find that, given our improved starting point, very few points are added or removed, but the heights and sizes of existing point are further updated.

For the second stage model, when sampling the auxiliary variable $\mathbf{v}$, we created a starting configuration of 3 randomly chosen location points per region and used a burn-in period of

2,000 iterations. To speed up sampling from the proposal density $q(v|\gamma_1, \gamma_2)$, we generated a fine-scaled grid of (γ_1, γ_2) values that span the support of the posterior

$p(\gamma_1, \gamma_2 | \mathbf{x}, \mu_{\gamma_1}, \tau_{\gamma_1}^2, \mu_{\gamma_2}, \tau_{\gamma_2}^2)$. Then we generated samples from $q(v|\gamma_1, \gamma_2)$ for each point on the (γ_1, γ_2) grid.

For our simulations, we let γ_1 and γ_2 each range between about 0.5 and 8, in increments of 0.2. We found that, for such fine-scaled grids, generating the samples across the grid could take a substantial amount time (for our implementation, it took a couple of days to run these simulations on an Intel Core i5 processor).

When implementing the sampling algorithm for the posterior of (γ_1, γ_2) , given in Appendix B.1, we use the configurations v that are sampled across the grid of (γ_1, γ_2) -values. The proposal density $q(\gamma_1, \gamma_2)$ that we use is defined on this (γ_1, γ_2) -grid. We give every point on the grid the same weight. This proposal density gives a 20% acceptance rate for the proposed (γ'_1, γ'_2) . We could achieve a higher acceptance rate by allowing the proposal to depend on the current (γ_1, γ_2) values, but we found that our proposal density has superior mixing properties.

Sampling the auxiliary variable v , we recognize that points may lie close to or on the boundaries of the brain region of interest. To get an accurate count of the number of neighbors of the boundary points, we created a boundary region for each region of interest. We design the boundary regions to be at least 5 voxels wide. Point processes v are then simulated on the expanded regions (original region plus boundary region).

As explained in Section 4.1, we implement our model for independent pairs of subjects. Given that our dataset includes 21 subjects, we have 210 unique pairs of subjects and a maximum 10 independent subject pairs. We fit our model repeatedly for several sets of 10 independent subject pairs. As an example, the subject indices corresponding to one such set of independent pairs could be: $S_{10}^* = \{(1, 2), (3, 4), \dots, (19, 20)\}$.

For a given set of independent subject pairs, we generate samples $\mu_{\gamma}^{(1)}, \dots, \mu_{\gamma}^{(L)}$ from the posterior. Then, assuming that we have K sets of subject pairs, we generate an MCMC chain $\mu_{\gamma}^{(k; 1:L)} = \{\mu_{\gamma}^{(k; 1)}, \dots, \mu_{\gamma}^{(k; L)}\}$ for each set, $k = 1, \dots, K$. To make inference about the between-

subject interaction, we calculate $g(\mu_{\gamma}) = \exp(\mu_{\gamma})$. Then, to estimate $E_{\pi} g = \int_{\Omega_{\mu}} g(\mu_{\gamma}) \pi(d\mu_{\gamma})$

where π is the target distribution (the posterior of μ_{γ}) and Ω_{μ} is the support of π , we calculate

$$\hat{g}^{(k)} = \frac{1}{L} \sum_{l=1}^L \exp(\mu_{\gamma}^{(k; l)})$$

and

$$\hat{g} = \frac{1}{K} \sum_{k=1}^K \hat{g}^{(k)}.$$

Similar to the method of batch means, we use the standard deviation of the $\hat{g}^{(k)}$ to estimate the Monte Carlo standard error of our estimate. For our model implementation, we use $K = 20$ sets of independent subject pairs and $L = 5000$ as the length of the MCMC chains.

Source code to reproduce the results is available as Supporting Information on the journals web page (<http://onlinelibrary.wiley.com/doi/xxx/supinfo>).

References

- Al-Awadhi F, Hurn M, Jennison C. Three-dimensional bayesian image analysis and confocal microscopy. *Journal of Applied Statistics*. 2011; 38(1):29–46.
- Baddeley A, Lieshout MV. Stochastic geometry models in high-level vision. *Journal of Applied Statistics*. 1993; 20(5-6):231–256.
- Baddeley A, Moyeed R, Howard C, Boyde A. Analysis of a three-dimensional point pattern with replication. *Applied Statistics*. 1993:641–668.
- Baddeley A, Turner R. Practical maximum pseudolikelihood for spatial point patterns. *Australian & New Zealand Journal of Statistics*. 2000; 42(3):283–322.
- Baddeley AJ, Van Lieshout M. Area-interaction point processes. *Annals of the Institute of Statistical Mathematics*. 1995; 47(4):601–619.
- Beckmann CF, Jenkinson M, Smith SM. General multilevel linear modeling for group analysis in fmri. *Neuroimage*. 2003; 20(2):1052–1063. [PubMed: 14568475]
- Bell ML, Grunwald GK. Mixed models for the analysis of replicated spatial point patterns. *Biostatistics*. 2004; 5(4):633–648. [PubMed: 15475424]
- Berthelsen, KK., Møller, J. *Case Studies in Spatial Point Process Modeling*. Springer; 2006. Bayesian analysis of markov point processes; p. 85-97.
- Bothma, A. Ph D thesis. University of Oxford; 2010. A model-based statistical approach to functional MRI group studies. URL <https://ora.ox.ac.uk/>.
- Brett M, Johnsrude IS, Owen AM. The problem of functional localization in the human brain. *Nature reviews neuroscience*. 2002; 3(3):243–249. [PubMed: 11994756]
- Diggle, PJ., Egle, SJ., Troy, JB. *Case Studies in Spatial Point Process Modeling*. Springer; 2006. Modelling the bivariate spatial distribution of amacrine cells; p. 215-233.
- Diggle PJ, Mateu J, Clough HE, et al. A comparison between parametric and non-parametric approaches to the analysis of replicated spatial point patterns. *Advances in Applied Probability*. 2000; 32(2):331–343.
- Dosenbach NU, Fair DA, Cohen AL, Schlaggar BL, Petersen SE. A dual-networks architecture of top-down control. *Trends in cognitive sciences*. 2008; 12(3):99–105. [PubMed: 18262825]
- Eckel S, Fleischer F, Grabarnik P, Kazda M, Sarkka A, Schmidt V. Modelling tree roots in mixed forest stands by inhomogeneous marked gibbs point processes. *Biometrical Journal*. 2009; 51(3):522–539. [PubMed: 19572316]
- Fan J, McCandliss BD, Fossella J, Flombaum JI, Posner MI. The activation of attentional networks. *NeuroImage*. 2005; 26(2):471–479. [PubMed: 15907304]
- Friston KJ, Glaser DE, Henson RN, Kiebel S, Phillips C, Ashburner J. Classical and bayesian inference in neuroimaging: applications. *Neuroimage*. 2002a; 16(2):484–512. [PubMed: 12030833]
- Friston KJ, Penny W, Phillips C, Kiebel S, Hinton G, Ashburner J. Classical and bayesian inference in neuroimaging: theory. *NeuroImage*. 2002b; 16(2):465–483. [PubMed: 12030832]

- Frost MA, Goebel R. Measuring structural–functional correspondence: spatial variability of specialised brain regions after macro-anatomical alignment. *Neuroimage*. 2012; 59(2):1369–1381. [PubMed: 21875671]
- Geyer, C. Likelihood inference for spatial point processes. In: Barndorff-Nielsen, O. Kendall, W., van Lieshout, M., editors. *Stochastic geometry: likelihood and computation*. Chapman & Hall; Boca Raton: 1999.
- Geyer CJ, Møller J. Simulation procedures and likelihood inference for spatial point processes. *Scandinavian journal of statistics*. 1994:359–373.
- Grabarnik P, Särkkä A. Interacting neighbour point processes: some models for clustering. *Journal of Statistical Computation and Simulation*. 2001; 68(2):103–125.
- Hartvig NV. A stochastic geometry model for functional magnetic resonance images. *Scandinavian Journal of Statistics*. 2002; 29(3):333–353.
- Heller R, Golland Y, Malach R, Benjamini Y. Conjunction group analysis: An alternative to mixed/random effect analysis. *Neuroimage*. 2007; 37(4):1178–1185. [PubMed: 17689266]
- Illian JB, Hendrichsen DK. Gibbs point process models with mixed effects. *Environmetrics*. 2010; 21(3–4):341–353.
- Jenkinson M, Beckmann CF, Behrens TE, Woolrich MW, Smith SM. FSL. *NeuroImage*. 2012; 62(2):782–790. [PubMed: 21979382]
- Kang J, Johnson TD, Nichols TE, Wager TD. Meta analysis of functional neuroimaging data via bayesian spatial point processes. *Journal of the American Statistical Association*. 2011; 106(493):124–134. [PubMed: 21706069]
- Kannurpatti SS, Motes MA, Rypma B, Biswal BB. Neural and vascular variability and the fmri-bold response in normal aging. *Magnetic resonance imaging*. 2010; 28(4):466–476. [PubMed: 20117893]
- Keller, M., Lavielle, M., Perrot, M., Roche, A. International Conference on Medical Image Computing and Computer-Assisted Intervention. Springer; 2009. Anatomically informed bayesian model selection for fmri group data analysis; p. 450–457.
- Kelly C, Toro R, Di Martino A, Cox CL, Bellec P, Castellanos FX, Milham MP. A convergent functional architecture of the insula emerges across imaging modalities. *Neuroimage*. 2012; 61(4):1129–1142. [PubMed: 22440648]
- Kim S, Smyth P, Stern H. A bayesian mixture approach to modeling spatial activation patterns in multisite fmri data. *Medical Imaging, IEEE Transactions on*. 2010; 29(6):1260–1274.
- Madhyastha TM, Askren MK, Boord P, Grabowski TJ. Dynamic connectivity at rest predicts attention task performance. *Brain connectivity*. 2015a; 5(1):45–59. [PubMed: 25014419]
- Madhyastha TM, Askren MK, Zhang J, Leverenz JB, Montine TJ, Grabowski TJ. Group comparison of spatiotemporal dynamics of intrinsic networks in parkinsons disease. *Brain*. 2015b:awv189.
- Møller, J. Markov chain monte carlo and spatial point processes. In: Barndorff-Nielsen, O. Kendall, W., van Lieshout, M., editors. *Stochastic geometry: likelihood and computation*. Chapman & Hall; Boca Raton: 1999.
- Møller, J., Waagepetersen, RP. *Statistical inference and simulation for spatial point processes*. CRC Press; 2003.
- Myllymäki M, Mrkvi ka T, Grabarnik P, Seijo H, Hahn U. Global envelope tests for spatial processes. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*. 2016
- Myllymäki M, Panoutsopoulou I, Särkkä A. Analysis of spatial structure of epidermal nerve entry point patterns based on replicated data. *Journal of microscopy*. 2012; 247(3):228–239. [PubMed: 22906010]
- Nelson SM, Dosenbach NU, Cohen AL, Wheeler ME, Schlaggar BL, Petersen SE. Role of the anterior insula in task-level control and focal attention. *Brain structure and function*. 2010; 214(5–6):669–680. [PubMed: 20512372]
- Nichols TE, Holmes AP. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human brain mapping*. 2002; 15(1):1–25. [PubMed: 11747097]
- R Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing; Vienna, Austria: 2015. URL <http://www.R-project.org/>.

- Ripley BD. Modelling spatial patterns. *Journal of the Royal Statistical Society Series B (Methodological)*. 1977;172–212.
- Rosenblatt JD, Vink M, Benjamini Y. Revisiting multi-subject random effects in fmri: Advocating prevalence estimation. *NeuroImage*. 2014; 84:113–121. [PubMed: 23988271]
- Schladitz K, Särkkä A, Pavenstadt I, Haferkamp O, Mattfeldt T. Statistical analysis of intramembranous particles using freeze fracture specimens. *Journal of microscopy*. 2003; 211(2): 137–153. [PubMed: 12887708]
- Thirion B, Flandin G, Pinel P, Roche A, Ciuciu P, Poline JB. Dealing with the short-comings of spatial normalization: Multi-subject parcellation of fmri datasets. *Human brain mapping*. 2006; 27(8): 678–693. [PubMed: 16281292]
- Thirion B., Varoquaux G., Poline JB. *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer; 2010. Accurate definition of brain regions position through the functional landmark approach; p. 241-248.
- Waller LA, Särkkä A, Olsbo V, Myllymaki M, Panoutsopoulou IG, Kennedy WR, Wendelschafer-Crabb G. Second-order spatial analysis of epidermal nerve fibers. *Statistics in Medicine*. 2011; 30(23):2827–2841. [PubMed: 21823143]
- Weeda WD, Waldorp LJ, Christoffels I, Huizenga HM. Activated region fitting: A robust high-power method for fmri analysis using parameterized regions of activation. *Human brain mapping*. 2009; 30(8):2595–2605. [PubMed: 19172652]
- Woolrich MW, Behrens TE, Beckmann CF, Jenkinson M, Smith SM. Multilevel linear modelling for fmri group analysis using bayesian inference. *Neuroimage*. 2004; 21(4):1732–1747. [PubMed: 15050594]
- Worsley KJ, Liao C, Aston J, Petre V, Duncan G, Morales F, Evans A. A general statistical analysis for fmri data. *Neuroimage*. 2002; 15(1):1–15. [PubMed: 11771969]
- Xu L, Johnson TD, Nichols TE, Nee DE. Modeling inter-subject variability in fmri activation location: A bayesian hierarchical spatial model. *Biometrics*. 2009; 65(4):1041–1051. [PubMed: 19210732]

Figure 1.
The activation surfaces (maps of z -statistics) of four of the subjects for the selected axial slice.

Figure 2.
Trace plots for the average height and size parameters.

Figure 3.

The activation surfaces (top row) and corresponding fitted surfaces $A_s = \{A_{s,v} | v \in V\}$ (bottom row) of four of the subjects for the selected axial slice.

Figure 4.

The four locations used for model checking superimposed on the average activation map

$$1/S \sum_{s=1}^S y_s$$

Figure 5.
The four regions of brain activation identified for the second stage analysis.

Figure 6.

The pooled cross-subject $\bar{g}(h)$ functions for Regions 1 through 4 (solid line). The dashed lines represent the pointwise 95% envelopes calculated under the assumption that the point patterns are independent across subjects.

Figure 7.

Trace plots of the posterior samples of $\exp(\mu_{\gamma_1})$, $\exp(\mu_{\gamma_2})$, $1/\tau_{\gamma_1}$ and $1/\tau_{\gamma_2}$ for one set of independent subject pairs, Region 2.

Figure 8.

The average cross-subject $\bar{L}(h) - h$ functions for each of the four regions (solid line). The dashed lines represent the 95% global rank envelopes calculated based on simulations from the fitted model.

Table 1

Summary measures for the configuration of activation units identified using the first stage model. The following subject-wise summary measures are calculated: number of units n_s , the average height $\bar{a}_s$ and size $\bar{d}_s$ of the activation units, the average mean square error (AMSE), and the maximum activation surface height $y_s^{\max}$. The means and standard deviations over all the subjects of these summaries are reported. For the height and size parameters, we also report the average within-subject standard deviation.

Measure	Mean	St. Dev. (between)	Mean St. Dev. (within)
n	137.33	34.95	
$\bar{a}_s$	3.10	0.52	1.52
$\bar{d}_s$	40.09	3.39	17.68
AMSE	0.23	0.07	
$y_s^{\max}$	9.80	3.19	

Table 2

Estimates of $\exp(\mu_{\gamma_1})$, $\exp(\mu_{\gamma_2})$, $1/\tau_{\gamma_1}$, $1/\tau_{\gamma_2}$ and the corresponding standard error estimates (in parentheses) for the four regions.

Region	$\exp(\mu_{\gamma_1})$	$\exp(\mu_{\gamma_2})$	$1/\tau_{\gamma_1}$	$1/\tau_{\gamma_2}$
Region 1 (left operculum and insular cortices)	1.52 (0.03)	1.91 (0.04)	0.35 (0.01)	0.47 (0.02)
Region 2 (right operculum and insular cortices)	1.42 (0.04)	2.11 (0.06)	0.37 (0.01)	0.55 (0.02)
Region 3 (left occipital cortex)	1.39 (0.03)	3.01 (0.07)	0.31 (0.01)	0.58 (0.03)
Region 4 (right occipital cortex)	1.38 (0.04)	3.43 (0.08)	0.35 (0.02)	0.63 (0.02)