BAYESIAN SPATIAL MODELING OF FMRI DATA: A MULTIPLE-SUBJECT ANALYSIS

by

Lei Xu

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2007

Doctoral Committee:

Research Associate Professor Timothy D. Johnson, Co-Chair
Adjunct Associate Professor Thomas E. Nichols, Co-Chair
Professor Douglas C. Noll
Assistant Professor Zhaohui Steve Qin

ACKNOWLEDGEMENTS

I wish to take this opportunity to extend my heartfelt thanks to my Co-chairs: Dr. Timothy Johnson and Dr. Tomas Nichols. They gave so unselfishly of their time and were tremendously helpful in this research work. They introduced this interesting problem to me, drew my attention to important papers, provided me with their insightful suggestions and diligently reviewed and edited my paper. Without their guidance, patience, encouragement, support and expertise in the field, this project could not have been completed.

I would also like to thank my other committee members, Dr. Douglas C. Noll and Dr. Zhaohui Steve Qin for their patience in reading the manuscript and valuable suggestions.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

LIST OF APPENDICES

Appendix

CHAPTER I

INTRODUCTION

The identification of a specific brain region with a specific function is a central problem in cognitive neuroscience. With the development of functional magnetic resonance imaging (fMRI), we are now able to follow the brain changes functionally during specific task performance.

It has been observed that activated areas of the human brain show localized increases in blood flow. These increases ensure an adequate supply of oxygen to regions working hard. Changes in the oxygenation level of the blood therefore occur as a consequence of neuronal activity. fMRI makes indirect use of this local need for energy and images the brain activity by visualizing regional changes in blood flow, blood volume and blood oxygenation. The most common fMRI technique utilizes blood oxygenation level dependent (BOLD) contrast, which is based on the differing magnetic properties of oxygenated and deoxygenated blood.

In a typical fMRI experiment, participants are given a fixed set of controlled stimuli and perform a task repeatedly. In many cases the stimulus is presented in a so-called boxcar design of alternating activation and rest periods. For example, a subject may view a picture for 30 seconds and rest for another 30 seconds, and so on for five cycles. During the experiment, a series of scans of the brain are taken

over time. Due to the delay of hemodynamic response, the BOLD signal increases gradually and typically reaches a peak 4-6 seconds after the onset of the task. Figure 1.1 shows a boxcar stimulus and the delayed version of the stimulus together with MR signal time series at voxels which are strongly, weakly and not activated. After the experiment has finished, the raw images require a series of spatial transformations to correct for distortions and reduce the effect of movement or shape differences among scans. Statistical methods are then used to measure activity at voxels through the amount of association between the stimulus and fMRI time series.

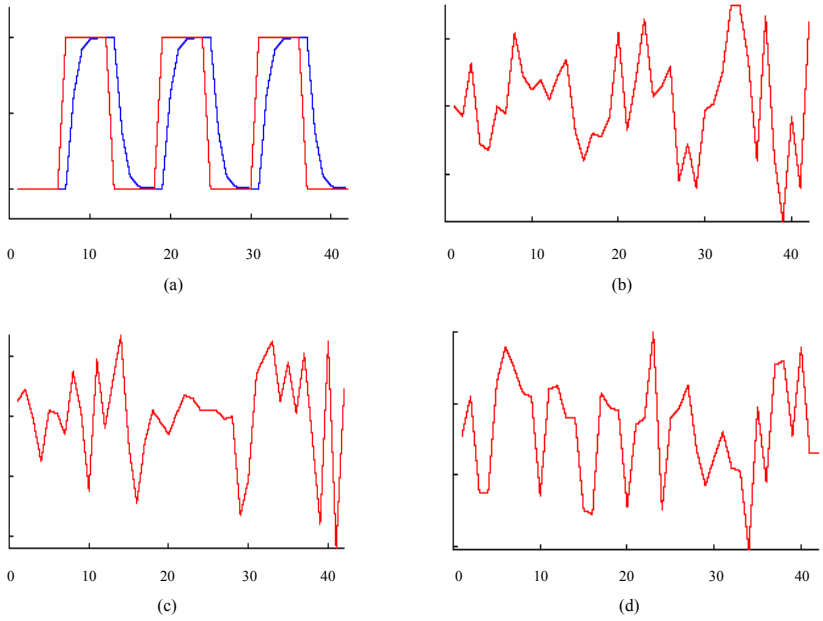


Figure 1.1: (a) the visual stimulus in red and the delayed stimulus in blue. (b), (c) and (d) the MR signals that are strongly, weekly and not activated

Our work is motivated by an event-related fMRI study (Park et al., 2003) in which the role of frontal and hippocampal structures in a working memory task was investigated. Twenty one adults participants were presented with two types of tasks. In the visual condition (top row of Figure 1.2), subjects viewed a scene for 6 seconds, followed by a picture fragment probe (4-second interval). They were to

press a key to indicate whether or not the probe fragment was part of the picture that they had just studied. Each trial was followed by a 12-second baseline period. In the maintenance condition (bottom row of Figure 1.2), subjects viewed a scene for 2 seconds, followed by a 4-second rehearsal interval during which they stored their memory of the pictures they had just seen. They then did a probe match (4-second interval), followed by a 12-second baseline. Subjects preformed 48 trials of each type presented in random order. Of particular interest was the 4-second interval where, in the maintenance condition, subjects were maintaining the stimulus but in the extended visual condition, they passively viewed the picture until the probe appeared. Comparison of performance relative to baseline in these conditions permitted them to determine what neural structures were used to perform each task.

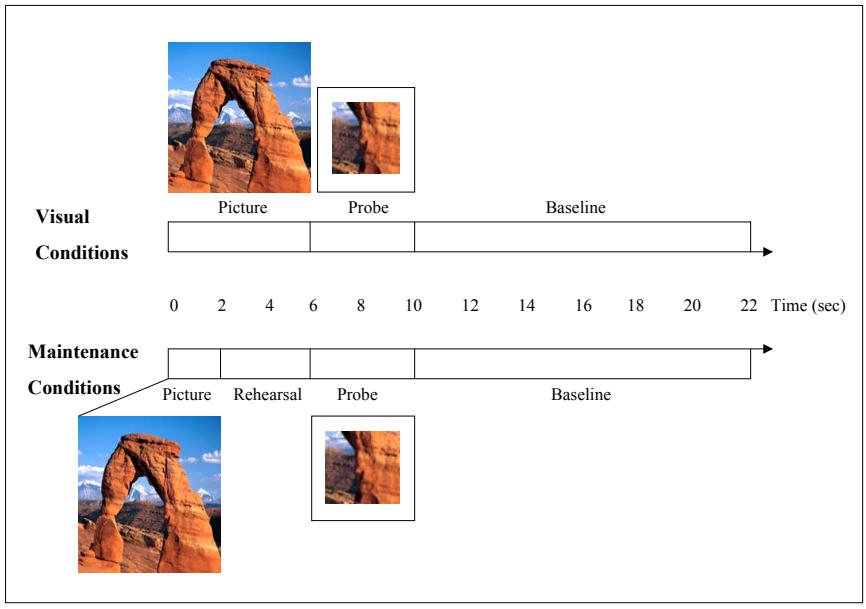


Figure 1.2: Working memory tasks for the visual and maintenance conditions

The spatial and temporal structure of the massive 4-D fMRI data imposes challenging problems in statistical modeling. There are two general approaches, classical and Bayesian. Classical analyses are typically two-stage (Friston et al., 1995). The

first stage is to fit a univariate time series model independently at each voxel and obtain a single image of some parameter estimates and test statistics such as t-values or z-scores. These statistics indicate evidence against the null hypothesis of no effect at each voxel. Let y_t be the observed signal intensity acquired at one voxel at time t and let $\mathbf{Y}$ be a vector of these observations. The most commonly used models are general linear models (GLM) of the following form: (Friston et al. 1995)

$$Y = X\beta + E \tag{1.1}$$

A very simple example is

$$y_t = \alpha + \beta x_t + e_t$$

where α is the baseline (rest) value and β is the effect parameter to be estimated. The regressors x_t is the delayed stimulus function which models the hemodynamic effects. A common choice for x_t is a convolution of a gamma or poisson density $h(t)$ with a delta function $\delta(t)$. $\delta(t)$ takes value of 1 or 0 indicating the presence or absence of the stimulus. The use of convolution allows the model to capture a range of different BOLD response shapes. The error term $\{e_t\}$ is often assumed to be normally distributed with mean 0 and variance σ^2. In multiple subject problems subject is treated as random effect so that inference can extend to the entire population. Analysis of the random effects is accomplished using what is referred to as the Summary Statistic Method. This involves first fitting a GLM of form (1.1) for each subject and then submitting the image of $c\hat{\beta}$ to a second-level analysis.

$$c\hat{\beta} = X_g\beta_g + E_g$$

The final t-images is based on the test statistics $t = \frac{c_g\hat{\beta}_g}{sd(c_g\hat{\beta}_g)}$, where c and c_g are the contrasts of interest.

The second stage is to threshold the t-values or z-scores to define regions of significant activation. One useful method is to perform a test at each voxel using a single threshold and look for voxels above the threshold. Various methods on how to choose the threshold have been proposed (Genovese, Lazar, and Nichols, 2002; Nichols and Hayasaka, 2003). As fMRI data contains thousands of voxels a large number of voxels will be significantly active simply by chance. This is the multiple comparisons problem. If tests are independent the overall type I error can be controlled using the Bonferroni correction. However, because of the spatial correlation inherent in fMRI data, the Bonferroni approach is often too conservative. Another method known as the Gaussian Random Field (GRF) Method has been developed recently to deal with the effect of spatial correlation. GRF method gives the threshold based on the expected number of significantly activated clusters. This method requires the image be sufficiently smooth to fulfill the assumption of a continuous random field.

While the classical method is a computationally efficient method, it has several limitations. The approach does not account for the spatial nature of the image data and does not utilize the prior knowledge that spatially contiguous patterns of activation are expected to have. Spatial dependence is either ignored or simply accounted for in the choice of threshold. Multisubject analyses are particularly problematic since, even after registration of the subjects' brain to a common atlas, there is residual variation in the anatomical landmarks; further, it has been shown that even if sulci and gyri are aligned there is variation in the functional landmarks (Morosan et. al., 2001). A mass-univariate model cannot account for any mismatch in activation location and will only detect voxels with consistent change in activation.

A lot of effort has been made on how to use the spatial information in the data to enhance signal detection. Some methods find a threshold that accounts for cor-

relation in the null hypothesis statistic images (Worsley et al., 1996; Friston et al., 1993). The most general approach is to spatially smooth the data with a fixed-width Gaussian kernel in a preprocessing step. However the Gaussian filter tends to blur and change the shape of the activation. Descombes, Kruggel, and von Cramon (1998) proposed an alternative to Gaussian filtering by restoring the signal using a spatiotemporal Markov Random Field (MRF). The MRF is used to define the prior knowledge of the underlying signal. This priori property is modeled by interactions between neighboring pixels and is incorporated in the procedure in a Bayesian framework. Polzehl and Spokoiny(2001) proposed a structural adaptive smoothing procedure, specifically the propagation-separation (PS) approach for time-series of images. Tabelow et al. (2006) provided a complete procedure for fMRI analysis using the PS approach and showed significant improvement on the information of the spatial extent and the shape of the activation region.

Many others proposed Bayesian methods to integrate the spatial modeling of the data in the statistical analysis. Hartvig et al. (2000) use a spatial mixture model and achieve computational feasibility by formulating the model through the marginal distribution on a small grid of voxels. Instead of using $p(\omega_i|y)$ they suggest to use $p(\omega_i|y_{c_i})$ and calculate it by summing over the neighboring states:

$$p(\omega_i|y_{c_i}) \propto \sum_{\omega_i^1=0}^{1} \cdots \sum_{\omega_i^m=0}^{1} f(y_{c_i}|\omega_{c_i})p(\omega_{c_i}) \tag{1.2}$$

where c_i is voxel i together with the m neighbors and $\omega_{c_i} = (\omega_i, \omega_i^1, \ldots, \omega_i^m)$ is the vector of all activation indicators in c_i. Hartvig's paper discusses three models for $p(\omega_{c_i})$ which models clustered activation based on s, the number of 1's in ω_{c_i}, and showes that the posterior probability is robust to the choice of models. The simplest one is:

$$p(\omega_{c_i}) = \begin{cases} q_0 & \text{if } s = 0, \\ q_1 & \text{if } s > 0. \end{cases}$$

The main contribution of their method is that by carefully choosing prior distributions of ω_{c_i} they are able to calculate the posterior distribution (1.2) analytically. This provides inference that is much faster than simulation based methods. In another paper Hartvig (2002) proposed a regression based spatial model using the idea of "activation centers". The model has the form of a GLM:

$$y_{it} = \beta_i x_t + \varepsilon_{it}$$

where $\beta_i \sim n(A_i(\theta), \tau^2)$ and x_t is the temporal variation caused by the BOLD effect. The activation surface $A_i(\theta)$ is modeled as a sum of Gaussian functions: $A_i(\theta) = \sum_{k=1}^{K} h(i, \theta_k)$, with parameters $\theta_k = (\mu_k, \alpha_k, r_k, \phi_k)$ describing the location, magnitude and shape of the k^{th} activation center. They design a reversible jump algorithm to insert, delete and change an activation center given the variance parameters are known. They estimate the variance parameters separately in a different procedure using the distribution of sufficient statistics. Cosman, Fisher, and Wells (2004) used an Ising MRF as the prior for neural activity. Woolrich et al. (2005)develop a spatial mixture model using a discrete Markov random field (MRF) priors on a spatial map of classification labels. For each observation y_i at voxel i, there is a binary latent class variable $\omega_i = 0$ or 1 denoting whether voxel i is active or inactive. Within this framework one can calculate the posterior probability of a voxel being activated:

$$p(\omega, \theta | y) \propto \{\prod_i f(y_i | \omega_i, \theta_{\omega_i})\} p(\omega) p(\theta) \tag{1.3}$$

where θ is the parameter in distribution f. The spatial regularization is controlled by a parameter ϕ in the discrete MRF which penalizes neighboring voxels of different

classes:

$$p(\omega|\phi) = f(\phi)\exp\{-\phi\sum_{i\sim j} I(\omega_i \neq \omega_j)\} \tag{1.4}$$

where $i \sim j$ implies i is a neighbourhood of j. $I(\omega_i \neq \omega_j)$ is an indicator function: $I(\omega_i \neq \omega_j) = 1$ if $\omega_i \neq \omega_j$ and zero otherwise. The problem with the discrete labels model is that the computational evaluation of the normalizing constant $f(\phi)$ is difficult. The spatial regularization parameter is typically assumed to be known. However, results can be highly dependent on this parameter. Hence, Woolrich et al. (2005) cleverly approximate the discrete MRF prior with a continuous MRF prior whose normalizing constant is easily calculated. They replaced the discrete labels with continuous weights π:

$$p(\pi, \theta|y) \propto \{\prod_i \sum_k \pi_{ik} f(y_i|\omega_i = k, \theta_k)\} p(\pi)p(\theta), \quad k = 0 \; or \; 1 \tag{1.5}$$

and define π using logistic transformation:

$$\pi_{ik} = \frac{\exp(\tilde{\pi}_{ik}/\gamma)}{\sum_k \exp(\tilde{\pi}_{ik}/\gamma)} \tag{1.6}$$

The prior of $\tilde{\pi}$ is now a continuous version of the MRF in (1.4):

$$p(\tilde{\pi}|\phi) = f(\phi)\exp\{-\phi\sum_{i\sim j}(\tilde{\pi}_{ik} - \tilde{\pi}_{jk})^2\}, \quad i = 1, ..., N \tag{1.7}$$

Because $\tilde{\pi}_{ik}$ is a continuous variable the normalizing constant $f(\phi)$ can be calculated easily: $f(\phi) \propto \frac{1}{\phi^N}$. Penny, Trujillo-Barreto, and Friston (2005) proposed a Gaussian MRF prior on the regression coefficient of a general linear model and approximate the posterior using Variational Bayes method. In a related paper Penny, Flandin, and Trujillo-Barreto (2007) show how the model evidence can be approximated using a Bayesian model comparison framework which allows the implementation of Analysis of Variance and Cluster of Interest analysis. Flandin and Penny (2007) replaced the

Gaussian MRF prior with s Sparse Spatial Basis Function prior which makes the use of a mixture model.

While these other methods model spatial dependence in the signal, they focus solely on single subject analyses. We are aware of only one method for multi-subject data. Miglioretti, McCulloch, and Zeger (2002) develop a multi-subject model for surface electrode data. Their model is developed for binary observations that is not applicable in the current setting.

In this work we propose a spatial model for multi-subject fMRI data in a hierarchical Bayesian framework. Our model improves the standard methods in several ways. First, we pose an explicit spatial model for activations at the subject level, unlike the mass-univariate method that fits each voxel independently of its neighbors. Second, our method is for multi-subject data and accounts for intersubject heterogeneity in activation location about a population location. Lastly, we use a fully Bayesian framework where all sources of uncertainty are considered and quantified. Throughout the paper we consider the effect magnitude image as our observations (or the data), one per subject, which may be interpreted roughly as map of percent change in blood flow, an indirect measure of neuronal activity. Although our model can be easily extended to analyze 3D data, in this paper we will only focus on one slice of the 3D image. To see an example, Figure 1.3 displays 5 of the 21 subjects from the working memory data set together with an anatomic reference image.

The dissertation is organized as follows. In Chapter II, we introduce a spatial mixture model in a full Bayesian framework. In Chapter III, we extend the prior class to nonparametric Dirichlet process. In Chapter IV, we conduct simulation studies and sensitivity analyses for both models. In Chapter V, we conclude the manuscript with a discussion and ideas for future work.

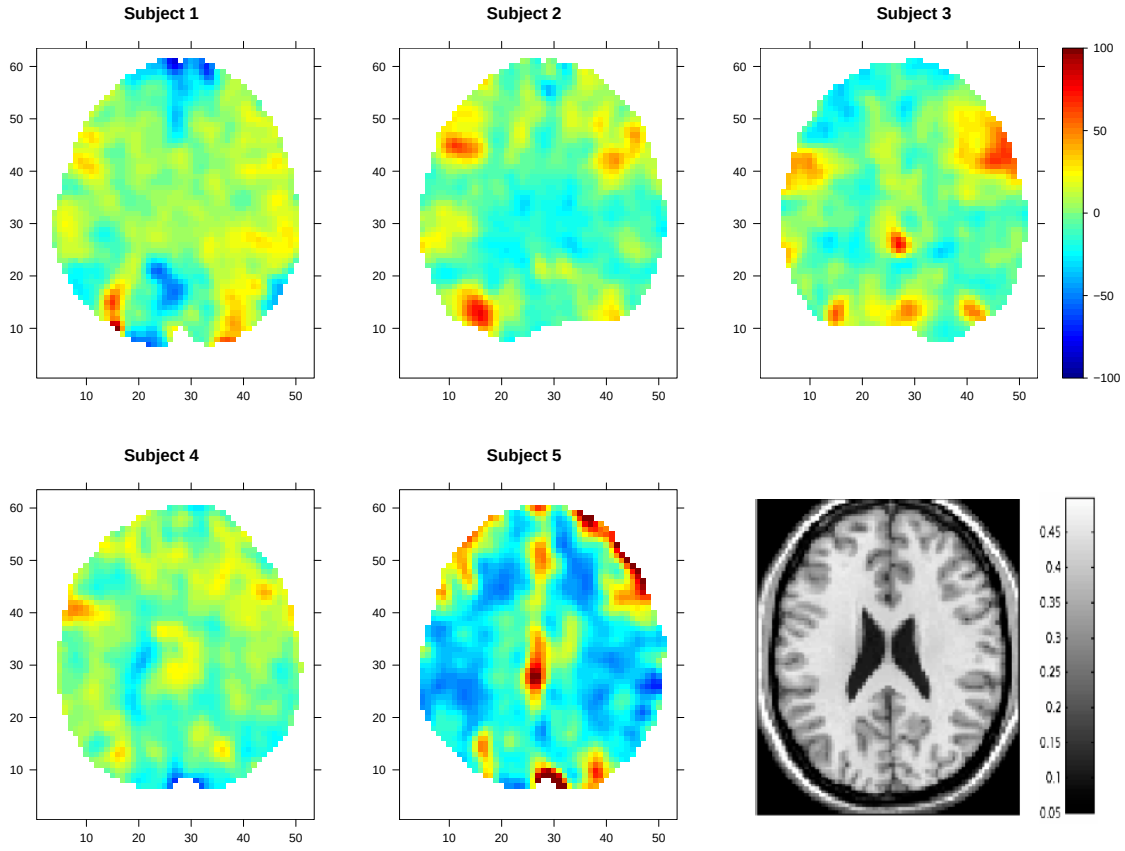


Figure 1.3: Contrast images (color) for subject 1 to 5 and an anatomic reference image in gray scale.

CHAPTER II

BAYESIAN SPATIAL MODELING OF FMRI DATA

2.1 Methods

We begin with an overview of the model and notation (with all subscripts initially suppressed), after which we present the distribution of the data and priors in detail.

Our model is specified hierarchically, as illustrated in Figure 2.1. At the first level we have an unknown number of "population centers", $\boldsymbol{\mu}$, that follow a homogenous spatial Poisson process defined over the confines of the brain. At the second level, an unknown number of "individual component" means, $\boldsymbol{\eta}$, are distributed as isotropic Gaussian mixtures whose mixture component means are the population centers with variances τ^2. We assume that each subject has been fit with an intrasubject fMRI signal model, producing scalar images of the fMRI blood oxygenation level dependent (BOLD) effect magnitude; we refer to these intrasubject summary measures as "the data". At the third level, we assume the data, y, for each subject are distributed as a Gaussian mixture with an unknown number of mixing components whose means are θ with variances σ^2. The mixing weights for the datum at pixel v are proportional to the density at v of a normal distribution with mean $\boldsymbol{\eta}$ and variance r^2. There is one special component representing the constant background intensity for each subject. We now present the details of our model.

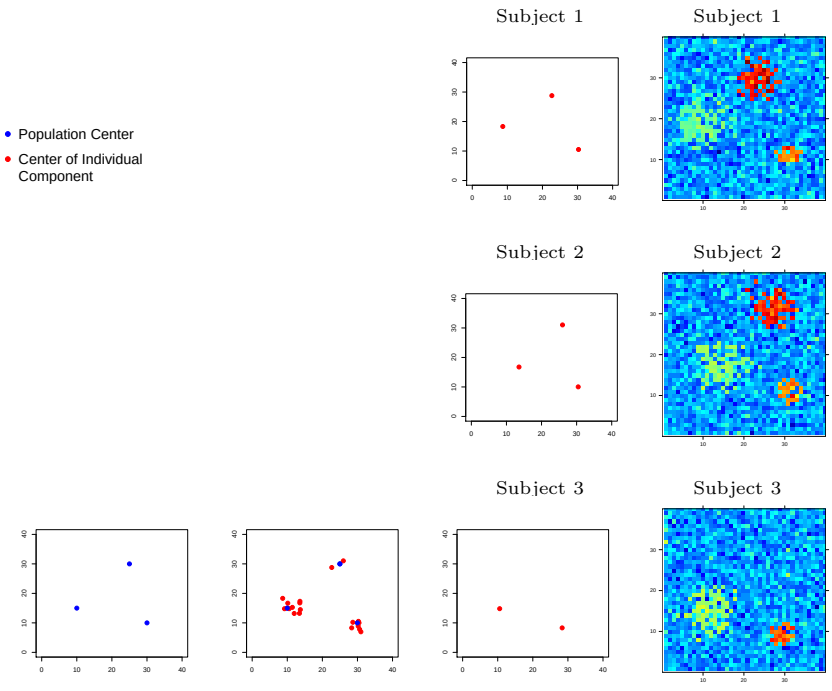


Figure 2.1: Hierarchical structure of the spatial Bayesian model. (a) At the first level an unknown number of "population centers", $\boldsymbol{\mu}$, that follow a homogenous spatial Poisson process defined over the confines of the brain. (b) At the second level, an unknown number of "individual component" means, $\boldsymbol{\eta}$, are distributed as isotropic Gaussian mixtures whose mixture component means are the population centers with variances τ^2. (c) At the final level, we assume the data, y, for each subject are distributed as a Gaussian mixture with an unknown number of mixing components whose means are θ with variances σ^2.

2.1.1 Data distribution

We begin with the third level of the hierarchy. Let y_{jv} be the observed effect magnitude at pixel v, $v = 1, \ldots, V$, for subject j, $j = 1, \ldots, J$. We use a single index v to reference a 2D pixel, $\mathbf{x}_v = (x_{1v}, x_{2v})$. We assume that observations are distributed as a mixture of Gaussian components:

$$f(y_{jv} \mid p_{jvl}, \theta_{jl}, \sigma^2_{jl}, c_j) = p_{jv0}\, \phi(y_{jv}; \theta_0, \sigma^2_0) + \sum_{l=1}^{c_j} p_{jvl}\, \phi(y_{jv}; \theta_{jl}, \sigma^2_{jl}). \tag{2.1}$$

Here, c_j is the number of mixture components for subject j (not including the background component) and $\phi(a; b, c^2)$ is the density at a of a normal distribution with mean b and variance c^2. θ_0 and σ^2_0 represent the mean and variance of the back-

ground component. By introducing a latent variable ω_{jv} with $\Pr(\omega_{jv} = l) = p_{jvl}$, the likelihood can be written conditionally as

$$\begin{aligned} f(y_{jv} \mid \theta_0, \sigma_0^2, \omega_{jv} = 0) &= \phi(y_{jv}; \theta_0, \sigma_0^2) \quad \text{for background,} \\ f(y_{jv} \mid \theta_{jl}, \sigma_{jl}^2, \omega_{jv} = l) &= \phi(y_{jv}; \theta_{jl}, \sigma_{jl}^2) \quad \text{otherwise.} \end{aligned}$$

The latent variable ω_{jv} can be interpreted as an allocation variable in the sense that observation y_{jv} is assigned to one of the mixture components with probability p_{jvl}. A pixel is considered active if that pixel belongs to a component with index $l > 0$ and is inactive if that pixel belongs to the background component ($l = 0$).

Let $\phi_2(\cdot; \mathbf{a}, \mathbf{B})$ denote the density of a bivariate normal distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{B}$. The mixing weights, p_{jvl}, take the form

$$p_{jvl} \propto \begin{cases} m & l = 0 \\ \phi_2(\mathbf{x}_v; \boldsymbol{\eta}_{jl}, \mathbf{R}_{jl}) & l = 1, \ldots, c_j \end{cases} \tag{2.2}$$

with $\sum_{l=0}^{c_j} p_{jvl} = 1$. Here $\mathbf{x}_v$ is the spatial location of pixel v and $\boldsymbol{\eta}_{jl}^{\mathrm{T}} = (\eta_{1jl}, \eta_{2jl})$ is the mean of individual component l, $l = 1, ..., c_j$ (we refer to the mean of the individual component as the "component center"). $\mathbf{R}_{jl} = r_{jl}^2 \mathbf{I}_2$ where $\mathbf{I}_2$ is the 2×2 identity matrix. The spatial dependence of the data is captured by the weights, p_{jvl}. Given m, the weights largely depend on the distance from the pixel to each of the component centers. This implies that observations that are spatially close to one another are more likely to have similar weights. Hence, spatially close observations will be more correlated than distant observations. Note that an observation distant from all components centers, a priori, will have probability near 1 that it belongs to the background center. Furthermore, if a pixel and component center are coincident, then the a priori probability that this pixel belongs to the background is approximately $m/(r_{jl}^{-2}/2\pi + m)$. fMRI experiments are usually designed such that only a

small percentage of the brain (roughly 1%–5%) actually shows any activity. Thus, a priori, we set $m = 19$ and $E(r_{jl}^{-2}) = (2\pi)$. This gives $m/(E(r_{jl}^{-2})/2\pi + m) = 0.95$, reflecting this a priori belief.

The number of mixing components, c_j, for each subject is not known and we estimate it along with all other parameters. A priori we assume c_j is a Poisson random variate with mean 5, for $j = 1, \ldots, J$.

Now we move on to the second and first levels of the hierarchy.

2.1.2 Priors

The priors of the component means make up the second level of the hierarchy. The prior of each component mean, $\boldsymbol{\eta}_{jl}$, $l = 1, \ldots, c_j$ for subject j, is taken to be a mixture of c_p bivariate normals. Each component mean is associated with a particular population center, $\boldsymbol{\mu}_i$, $i = 1, \ldots, c_p$:

$$\pi(\boldsymbol{\eta}_{jl} \mid \psi_i, \boldsymbol{\mu}_i, \tau_i^2, c_p) = \sum_{i=1}^{c_p} \psi_i \, \phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \tau_i^2 \mathbf{I}_2). \tag{2.3}$$

Here, the ψ_i are mixing weights. We also introduce a latent variable, z_{jl}, such that $\Pr(z_{jl} = i) = \psi_i$. Thus, conditional on $z_{jl} = i$, $\boldsymbol{\eta}_{jl}$ is bivariate normal with mean $\boldsymbol{\mu}_i$ and covariance matrix $\tau_i^2 \mathbf{I}_2$. A natural choice for the prior on ψ_i is a symmetric Dirichlet distribution: $\psi_i \mid c_p \sim D(1, 1, \ldots, 1)$ where the parameter dimension is $c_p - 1$. The number of population centers, c_p, is a priori unspecified and is to be estimated along with all other model parameters.

At the first level of the hierarchy, the parameter $\boldsymbol{\mu}_i^{\mathrm{T}} = (\mu_{1i}, \mu_{2i})$ is the location of population activation center i, $i = 1, ..., c_p$. Let A_j denote the cross-sectional area of the given MRI slice of the brain of subject j, $j = 1, \ldots, J$. Set $A = \cup_{j=1}^{J} A_j$. (We note here that, although all subject's data have been mapped onto a common brain atlas, due to motion artifacts and field inhomogeneities, there are missing data. Typically,

fMRI analyses are performed on the intersection of the A_j. By taking the union, we allow for the possibility that a population center is in a region where some subjects may have missing data.) A priori, we assume these population activation centers follow a homogenous spatial Poisson process with rate λ_p defined over A. Thus, we can write

$$c_p \sim P(\lambda_p A) \quad \text{and} \quad \pi(\boldsymbol{\mu}_1, \ldots, \boldsymbol{\mu}_{c_p} \mid c_p) = A^{-c_p}.$$

We take $\lambda_p = 5A^{-1}$ reflecting our belief that the number of population centers should be small. Note that although the prior distribution of the population centers is a homogenous point process, it's posterior is not necessarily. The posterior will depend, in large part, on the posterior distribution of the $\boldsymbol{\eta}_{jl}$.

We place inverse gamma (parameterized such that if $x \sim IG(\alpha, \beta)$, $E(x) = \beta/(\alpha - 1)$) priors on all variance parameters for mathematical convenience:

$$r_{jl}^2 \sim IG(2\pi, \beta_r),\ \tau_i^2 \sim IG(3, \beta_\tau),\ \sigma_{jl}^2 \sim IG(3, \beta_\sigma),\ \sigma_0^2 \sim IG(.001, .001).$$

Hyperprior distributions are then placed on β_r, β_τ and β_σ: $\beta_\sigma \sim Ga(.01, .01)$, $\beta_\tau \sim Ga(.01, .01)$, $\beta_r \sim Ga(2, 1)$ where $Ga(\kappa, \lambda) = \frac{\lambda^\kappa}{\Gamma(\kappa)} x^{\kappa-1} e^{-\lambda x}$ is the gamma distribution. We place rather vague hyperprior distributions on the scale parameters β_τ and β_σ to reduce the influence of the prior on the posterior estimates of τ_i^2 and σ_{jl}^2.

It remains to specify the priors of the intensity parameters, θ_{jl}, in the likelihood specified in equation (2.1). Recall that θ_{jl}, $l > 0$, are the mean intensity levels of the individual mixture components for active pixels and θ_0 is the mean intensity level of inactivated pixels. We choose the following priors:

$$\theta_0 \sim N(0, 1) \quad \text{and} \quad \theta_{jl} \sim \text{trunc}_{(0,\infty)} N(\lambda_\theta, \sigma_\theta^2),$$

where

$$\lambda_\theta \sim N(35, 1 \times 10^8) \quad \text{and} \quad \sigma_\theta^2 \sim IG(.01, .01).$$

Furthermore, $\text{trunc}_{(0,\infty)}N(\lambda_\theta, \sigma_\theta^2)$ denotes the normal distribution with mean λ_θ and variance σ_θ^2 truncated to $(0, \infty)$.

Results from simulation and sensitivity analyses are given in Chapter V.

2.1.3 Posterior Distribution

By combining the data distribution and the prior distributions the posterior distribution is

$$
\begin{aligned}
& \pi\left(\{\theta_{jl}\},\, \{\boldsymbol{\eta}_{jl}\},\, \{\boldsymbol{\mu}_i\},\, \{\psi_i\},\, \{\sigma^2_{jl}\},\, \{r^2_{jl}\},\, \{\tau^2_i\},\, \{c_j\},\, c_p,\, \beta_r,\, \beta_\tau,\, \beta_\sigma,\, \lambda_\theta,\, \sigma^2_\theta \mid \mathbf{y}\right) \\
\propto\; & f\left(\mathbf{y} \mid \{\theta_{jl}\},\, \{\boldsymbol{\eta}_{jl}\},\, \{\sigma^2_{jl}\},\, \{r^2_{jl}\},\, \{c_j\}\right) \pi\left(\{\theta_{jl}\} \mid \{c_j\}\right) \\
& \pi\left(\{\boldsymbol{\eta}_{jl}\} \mid \{\psi_i\},\, \{\boldsymbol{\mu}_i\},\, \{\tau^2_i\},\, \{c_j\}\right) \times \\
& \pi\left(\{\psi_i\} \mid \lambda_\psi,\, c_p\right) \pi\left(\{\boldsymbol{\mu}_i\} \mid c_p\right) \pi\left(\{\sigma^2_{jl}\} \mid \beta_\sigma,\, \{c_j\}\right) \pi\left(\sigma^2_0 \mid \beta_{\sigma_0}\right) \pi\left(\{\tau^2_i\} \mid \beta_\tau,\, c_p\right) \times \\
& \pi\left(\{r^2_{jl}\} \mid \alpha_r,\, \beta_r,\, c_p\right) \pi\left(\{c_j\}\right) \pi\left(c_p\right) \pi\left(\beta_r\right) \pi\left(\beta_\tau\right) \pi\left(\beta_\sigma\right) \pi\left(\lambda_\theta\right) \pi\left(\sigma^2_\theta\right) \\
= \;& \prod_{j=1}^{n} \prod_{v=1}^{V} \left(p_{jv0} (2\pi\sigma^2_0)^{-1/2} \exp\left[-0.5(y_{jv} - \theta_0)^2/\sigma^2_0\right] + \right. \\
& \left. \sum_{l=1}^{c_j} p_{jvl} (2\pi\sigma^2_{jl})^{-1/2} \exp\left[-0.5(y_{jv} - \theta_{jl})^2/\sigma^2_{jl}\right] \right) \times \\
& (2\pi)^{-1/2} \exp\left(-0.5\theta^2_0\right) \prod_{j=1}^{n} \prod_{l=1}^{c_j} (2\pi\sigma^2_\theta)^{-1/2} \exp\left[-0.5(\theta_{jl} - \lambda_\theta)^2/\sigma^2_\theta\right] \; I(\theta_{jl} > 0) \times \\
& \prod_{j=1}^{n} \prod_{l=1}^{c_j} \sum_{i=1}^{c_p} \psi_i (2\pi\tau^2_i)^{-1} \exp\left[-0.5(\boldsymbol{\eta}_{jl} - \boldsymbol{\mu}_i)^T (\boldsymbol{\eta}_{jl} - \boldsymbol{\mu}_i)/\tau^2_i\right] \times \\
& \frac{\Gamma(c_p \lambda_\psi)}{\Gamma(\lambda_\psi)^{c_p}} \prod_{i=1}^{c_p} \psi_i^{\lambda_\psi - 1} \prod_{i=1}^{c_p} \left[\frac{\beta^3_\tau}{\Gamma(3)} (\tau^2_i)^{-3-1} \exp\left(-\beta_\tau/\tau^2_i\right) \right] \times \\
& \prod_{j=1}^{n} \prod_{l=1}^{c_j} \left[\frac{\beta^{2\pi}_r}{\Gamma(2\pi)} (r^2_{jl})^{-2\pi-1} \exp\left(-\beta_r/r^2_{jl}\right) \; \frac{\beta^3_\sigma}{\Gamma(3)} (\sigma^2_{jl})^{-3-1} \exp\left(-\beta_\sigma/\sigma^2_{jl}\right) \right] \times \\
& \frac{0.001^{0.001}}{\Gamma(10^{-3})} (\sigma^2_0)^{-10^{-3}-1} \exp\left(-10^{-3}/\sigma^2_0\right) \frac{1}{\Gamma(2)} \beta^{2-1}_r \exp\left(-\beta_r\right) \times \\
& \frac{0.01^{0.01}}{\Gamma(10^{-2})} \beta^{10^{-2}-1}_\tau \exp\left(-10^{-2}\beta_\tau\right) \frac{0.01^{0.01}}{\Gamma(10^{-2})} \beta^{10^{-2}-1}_\sigma \exp\left(-10^{-2}\beta_\sigma\right) \times \\
& (2\pi 10^8)^{-1/2} \exp\left(-0.5(\lambda_\theta - 35)^2/10^8\right) \frac{0.01^{0.01}}{\Gamma(10^{-2})} (\sigma^2_\theta)^{-10^{-2}-1} \exp\left(-10^{-2}/\sigma^2_\theta\right) \times \\
& [5^{c_p} \exp(-5)/c_p!] \prod_{i=1}^{c_p} I(\boldsymbol{\mu}_i \in \cup_{j=1}^{n} A_j) \prod_{j=1}^{n} \left[5^{c_j} \exp(-5^{c_j})/c_j!\right].
\end{aligned}
$$

2.2 Posterior estimation

The full posterior distribution does not have an analytic solution. Thus, the posterior distribution is simulated via Markov chain Monte Carlo (MCMC) techniques, and, in particular reversible jump MCMC (RJMCMC). RJMCMC was first introduced by Green (1995) and can be viewed as a Metropolis-Hastings (MH) (Hastings, 1970) algorithm adapted to varying dimensional parameter spaces. In our example this corresponds to the addition and deletion of a population center or an individual component.

Briefly, we propose to add a new population center or an individual component, each with probability 0.5 and propose to delete a population center or individual component, each with probability 0.5 at each iteration of the algorithm. We oversample the RJMCMC moves three times per iteration which results in better mixing. When we propose to add an individual component, the parameters defining the new component are drawn from their prior distributions. The use of the prior distributions in proposing new values leads to a simplification of the acceptance probability. When we propose to add a population center, a new location $\boldsymbol{\mu}_*$ and variance τ_*^2 are drawn from their respective priors. A new mixing weight ψ_* is drawn from Beta$(1, c_p)$ and its kernel cancels the Jacobian of the transformation. We re-scale the old weights $\psi' s$ according to $\psi' = \psi(1 - \psi_*)$ such that all weights sum up to 1. The deletion move is the inverse of this construction. Conditional on the number of population centers and individual components other parameters are updated using a Gibbs or a random walk MH step. The variances in the proposal distribution for the MH steps were chosen to obtain acceptance rates of approximately 35%. Following Fernandez and Green (2002) we use the marginal expression for the likelihood and for the priors

of the $\boldsymbol{\eta}_{jl}$, as specified in equations (2.1) and (3.3) to obtain better mixing. Details are given in the Appendix A.

2.3 Summarizing posterior inference

Simultaneous visualization of the joint posterior distribution of all parameters is infeasible. Instead we view the distributions of certain univariate parameters and create images summarizing the posteriors of the various spatial parameters. In this section we review the approaches we have used to understand our posterior and assess model fit.

We create a "Posterior Probability of Activation" image for each subject, the pixel-wise posterior probability that a pixel is activated. Precisely, we estimate the marginal posterior probability that subject j has pixel v activated, $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. The Bayesian estimate of this quantity is the marginal posterior mean of $I_{(\omega_{jv}>0)}$, where $I_{(\omega_{jv}>0)} = 0$ if $\omega_{jv} = 0$ and 1 otherwise.

We create an image of the average "Individual Component Posterior", a pixelization of the (2D) posterior predictive density of the location of a new individual component for a new subject. Precisely, we seek to estimate the posterior distribution of $\tilde{\boldsymbol{\eta}}$, the location of a single individual component for a randomly selected subject. At each sweep we compute the marginal distribution (over population centers and subjects) of the individual component locations, $\sum_{i=1}^{c_p} \psi_i \, \phi_2(\mathbf{x}_v; \boldsymbol{\mu}_i, \tau_i^2 \mathbf{I}_2)$ for each v. Averaging this over sweeps creates an estimate of the density of $\tilde{\boldsymbol{\eta}} \mid \boldsymbol{y}$.

The Individual Component Posterior shows the most probable locations of individual components $\boldsymbol{\eta}$ and is most valuable for visualizing the spread of individual component centers about the population centers, as parameterized by τ_i^2. Care must be taken not to over-interpret the relative height of this image, however. As can be

seen by the prior for $\boldsymbol{\eta}$ (equation (3.3)), the mode height in the Individual Component Posterior are affected by three things. First, all things equal, a smaller τ_i^2 means a more concentrated density which will result in a higher mode. Second, a mode's height will grow as more subjects have components associated with that population center; e.g. if component l is associated with center i, then as $\sum_j I_{(z_{jl}=i)}$ grows so will ψ_i, the mixing weight for center i. And lastly, a subject that requires more components to fit their profile of activation ($\{y_{jv}\}_v$) associated with center i will also raise ψ_i.

We also create four images to characterize the population centers. The "Population Center Location" image is a pixelization of the (2D) posterior rate function (counts per pixel) for $\boldsymbol{\mu}$. We estimate this by computing the 2D histogram of $\{\boldsymbol{\mu}_i\}$ for each iteration and then average this over iterations. Note that this image does not sum to one, but rather sums to the posterior mean number of population centers, $E(c_p \mid \boldsymbol{y})$.

The "Population Center Scale" image shows the standard deviation of individual component locations about population centers. Exactly, this the posterior average of τ conditional on a population center being in pixel v. Note here we have intentionally suppressed the population center index i, as we marginalize over population center.

The "Population Center Prevalence" image shows the fraction of subjects that possess a population center. To precisely define this quantity, let $e_{ij} = \sum_l I_{(z_{jl}=i)}$ be the count of individual components that subject j has for population center i. The count of subjects with center i is then $e_i = \sum_{j=1}^{J} I_{(e_{ij}>0)}$, and the population prevalence is e_i/J. Conditional on a population center being in pixel v, this image is the posterior average of e_i/J after marginalizing over population centers $i = 1, ..., c_p$.

The "Population Center Average Area" image shows the average number of ac-

tivated pixels for a population center. Let D_i be the set of individual components that associated with population center i. For each individual component jl in D_i the number of activated pixels labeled with that component is $\sum_v I_{(\omega_{jv}=l)}$. Conditional on a population center being in pixel v, this image at v is the posterior average of $\sum_{jl \in D_i} \sum_v I_{(\omega_{jv}=l)}/e_i$, where e_i is the count of subjects with population center i defined in the previous paragraph.

For comparison with other methods we create two other images. A "Classical t-image" is the one-sample t-test on the BOLD effect magnitudes ($\boldsymbol{y_v} = \{y_{jv}\}_j$) at each voxel. To make inference on this result, we compute an image of $-\log_{10}$P-values for a one-sided, one sample t-test. The $-\log_{10}$ transformation makes for easier visualization, creating an image that should be "bright" in pixels with evidence for an non-null effect magnitude.

2.4 Real Data Results

We now summarize results of our model on the working memory data. We run the algorithm for 10,000 iterations with a burn-in of 5,000, saving every 5th iteration to summarize results. The algorithm takes approximately 105 minutes to run on a Mac 2.7 GHz PowerPC G5 processor. The algorithm was written in C++. The acceptance rate for the population level birth/death RJMCMC is about 5% and about 10% for the individual level . The number of population centers, c_p, converges at about 4000 iterations. The number of individual component centers, c_j, converges at about 1000 iterations. Trace plots and the posterior distribution of c_p and c_j for all 21 subjects are included in Appendix C. Figure 2.2 shows source data (y), the Posterior Probability of Activation and the location of individual component centers at iteration 6000 for a representative sample of subjects (1, 3, 9 and 20).

The Posterior Probability of Activation images demonstrate our model capturing the focal signals in subjects 1, 3 and 9, but appropriately finding no signal in subject 20. Even though subject 20 doesn't appear to have any activated pixels, Figure 2.2 (bottom row) still show four individual component centers for subject 20 at iteration 6000. This is due to the randomness of the RJMCMC step. Individual component centers can be proposed and accepted in a birth step for inactivated areas. However, these centers will die soon after. Only consistently activated pixels will be kept and shown on the Posterior Probability images. At some activation areas, our model fits several individual component centers instead of one, e.g. the upper right activation area for subject 2. This issue will be discussed in Chapter IV, the Future Work and Discussion. Results for all 21 subjects can be found in Appendix C.

The main focus of our work is the population level, as shown in Figure 2.3(a,b,c,d,e). Figure 2.3(a) displays the Population Center Location posterior; this image shows evidence of 5 to 7 population centers, localized to a a few pixels. Figure 2.3(b) shows the Population Center Prevalence image; of the 5 to 7 centers, 4 are present in more than 50% of the subjects: The bottom left and right bilateral pair of activations are in the dorsal occipital lobe, an area involved in visual processing; the upper left activation is in the dorsolateral prefrontal cortex (DLPFC), an area important maintaining "on-line", short-term representations of information. Figure 2.3(c) shows the Population Center Scale image, which quantifies the spread of individual components in terms of τ. The dispersal of individual components is mostly homogeneous ($\tau \approx 2$), with only the relatively rare right frontal center having larger spread. Figure 2.3(d) shows the Individual Component Posterior and gives similar information about the location of the 3 population centers as does Figure 2.3(a). However, note that the variability of the population locations is much smaller than the variability across

subjects. While Figure 2.3(a) gives more precise information about the population center location, Figure 2.3(d) qualitatively depicts the spread of individual components about population centers. Figure 2.3(e) shows the Population Activation Area image, which gives the average number of activated pixels for a population center. Most of the population activation areas consist of 6-10 pixels.

Figure 2.3 (f,g) display the classical t-image and the $-\log_{10}(p)$ image. While these images similarly illustrate the spread of activation between subjects, they cannot quantify the spatial precision of the results as our method does.

Our result is both more informative and less noisy. First, our method separately provides inference on (1) location of population activations, (2) the consistency with which subjects express such activations, and (3) the intrasubject spread of activations about the population centers. The classical method can only assess consistency of signal change at each voxel. With no spatial model, it cannot effect the smoothing of our model, nor can it account for intersubject variability in activation location.

The main contributions of this work are as follows. A traditional voxel-based approach can only detect activated voxels for which the average signal is significant across subjects. However, subjects sometimes don't have, or only a small percentage of the total subjects have, overlapped activated voxels even though these activated voxels are close to each other and represent the same anatomical/functional area. Our models fit individual component centers for subject level activation areas and find population activation centers based on these individual component centers. We explicitly model the location of the population activation center with a separate parameter describing the variability of the individual component centers. Unlike the voxel-based approach our methods allow for variation in both signal intensity and activation location at the individual level. The location of the population activa-

tion centers are summarized using the posterior rate function. This provides better precision in the estimation of the population center location than the standard mass-univariate method.

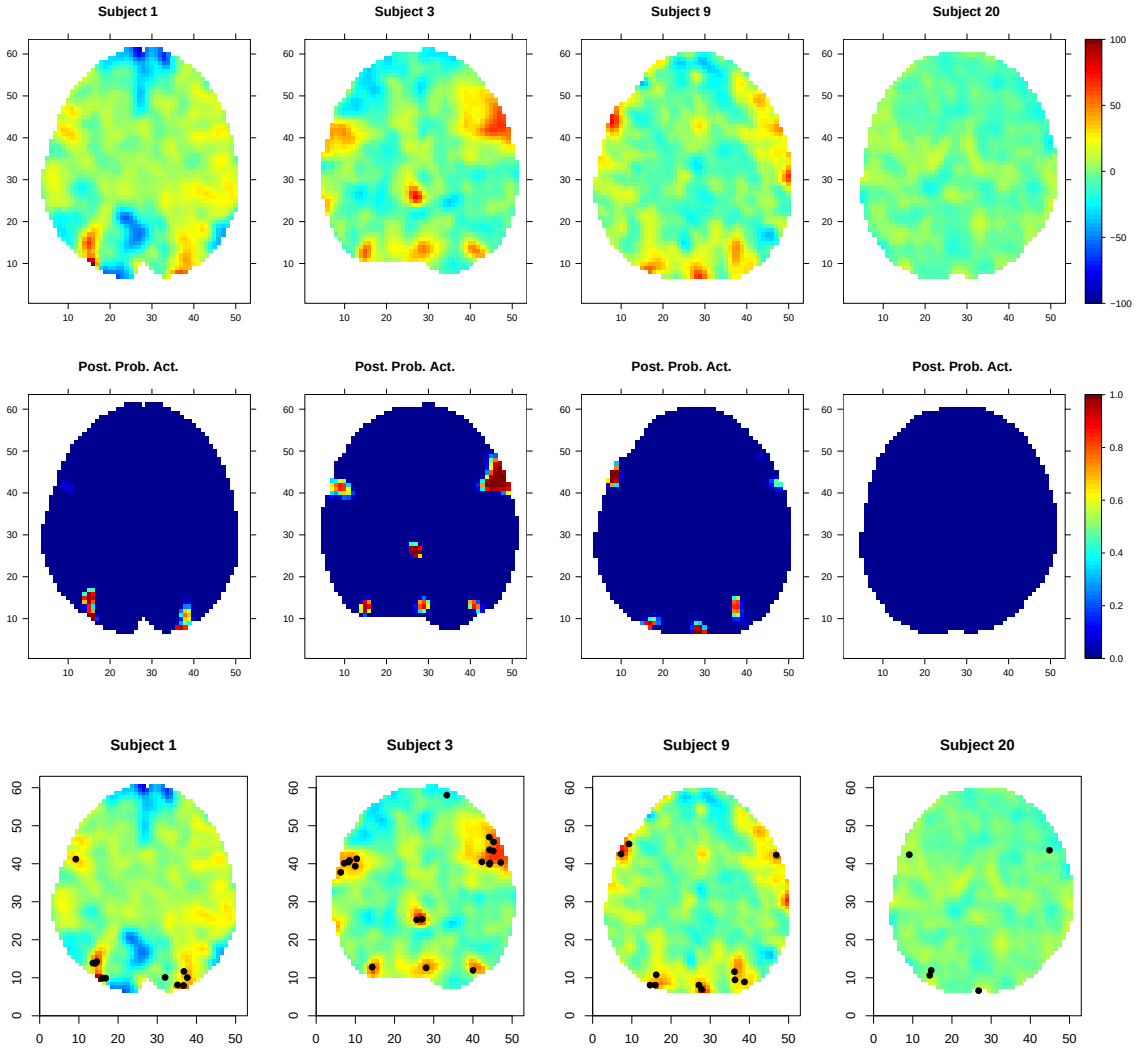


Figure 2.2: Top row: The intensity data from slice 21 for 4 subjects. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

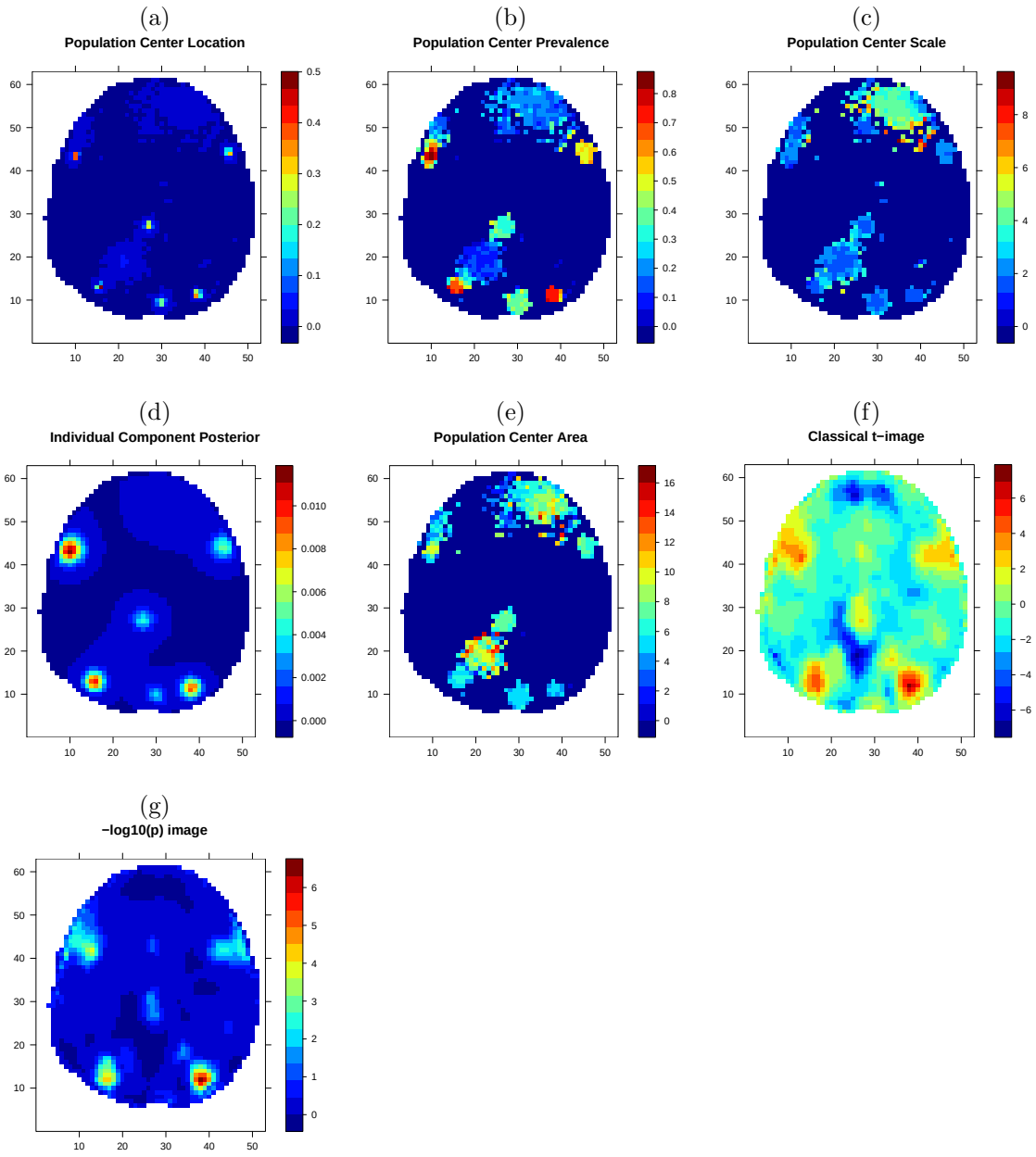


Figure 2.3: (a) 2D posterior rate function of population center locations; (b) proportion of subjects containing evidence for a population center; (c) standard deviation of individual components about population centers; (d) posterior-predictive density of individual center locations $p(\tilde{\boldsymbol{\eta}} \mid \mathbf{y})$; (e) the average number of activated pixels for a population center (f) classical t-image; (g) minus log base 10 p-values from the t-image $(-\log_{10}(p))$.

CHAPTER III

NONPARAMETRIC BAYESIAN MODEL USING DIRICHLET PROCESS PRIOR

3.1 Introduction

In Chapter II we introduce a spatial mixture mode (the "finite mixture" model). We assume each individual component center $\boldsymbol{\eta}_{jl}$ is from a mixture of c_p bivariate normals with mean $\boldsymbol{\mu}_i$ and covariance $\begin{bmatrix} \tau_i^2 & 0 \\ 0 & \tau_i^2 \end{bmatrix}$ for $i = 1, \ldots, c_p$. We interpret the normal mean $\boldsymbol{\mu}_i$ as the population center and estimate c_p using reversible jump MCMC. In this chapter, we extend the finite mixture model in two ways. First we extend the finite mixture model to an infinite mixture model. It can be shown that the infinite mixture model is equivalent to a nonparametric model assuming a Dirichlet process prior on the mean and covariance of the normal mixtures (Neal, 2002). In this setting the number of population centers c_p will be determined through the Dirichlet process. Second, we use a general variance-covariance matrix Σ_i for the bivariate normals. This helps capture different shapes of activations.

We first give a brief review of the Dirichlet process prior. We then describe in detail how the Dirichlet process prior is used to model the population and individual activations and how it is incorporated in the Bayesian hierarchical structure.

3.2 Review of Dirichlet Process Prior

In a fundamental paper on Bayesian nonparametric approaches, Ferguson (1973) defined the Dirichlet process as a random probability measure and derived many important properties of this process. Let $A_1, ..., A_k$ be a partition of the sample space S. We say a random probability measure G is a Dirichlet process if $G(A_1), ..., G(A_k)$ follows a Dirichlet distribution with parameter $\alpha(A_1), ..., \alpha(A_k)$. In most applications the nonnegative measure α is expressed as $\alpha(A) = \alpha_0 G_0(A)$ where $G_0(A)$ is a probability measure and α_0 is a positive scalar. In a Bayesian framework G_0 is the best prior guess of G and is the mean distribution of the Dirichlet process, α_0 is a precision parameter measuring the strength in the belief that G is G_0. Let y_i be the data and θ_i be the parameters. The basic Bayesian model using Dirichlet processes (DP) prior is often in the following form:

$$y_i \sim f(y_i|\theta_i),\ \theta_i \sim G,\ G \sim DP(\alpha_0 G_0) \tag{3.1}$$

The next important issue is to be able to simulate from the posterior distribution. Antoniak (1974) showed that the posterior distribution of the parameter θ given the data y is a mixture of Dirichlet processes. In practice a mixture of Dirichlet processes is difficult to use. Recently many authors have been working on solving the computational difficulties. Escobar (1994) develops a Gibbs sampler algorithm for f being the normal distribution with mean θ_i and variance 1. The conditional distribution of θ_i given all other $\theta_j's$, $(j \neq i)$ and the data y can be sampled according to the following rule:

$$\theta_i|\theta_j, j \neq i, y \begin{cases} = \theta_j & \text{with probability} \propto f(y_i|\theta_j), \\ \sim h(\theta_i|y_i) & \text{with probability} \propto A(y_i). \end{cases} \tag{3.2}$$

where $A(y_i) = \alpha_0 \int f(y_i|\theta_i) G_0(d\theta_i)$ and $h(\theta_i|y_i) = \frac{\alpha_0}{A(y_i)} f(y_i|\theta_i) g_0(\theta_i)$, g_0 is the density corresponding to G_0. Escobar and West (1995) extended the computational method to a more general likelihood $y|\theta \sim N(\mu, V)$ with a conjugate normal/inverse gamma prior for $G_0(\mu, V)$. Muller, Erkanli, and West (1996) applied the Dirichlet process prior to density estimation problems and also extended the model to multi-dimensional versions. Their sampling method is based on updating ϑ and s, where $\vartheta = \{\vartheta_1, ..., \vartheta_k\}$ is the set of distinct $\theta_i' s$ and s_i is a latent class variable associated with observation y_i such that $y_i \sim f(y_i|\vartheta_{s_i})$. This algorithm involves inserting a new component and deleting an empty component of ϕ. Both of these papers are based on conjugate priors and require an approximate draw from $p(\theta|y)$ by approximating $A(y)$. Later a new method is proposed by MacEachern and Muller (1998) which not only extend the model to non-conjugate situations but also avoid the difficult evaluation of $A(y)$ and gives an exact sampling method. They called it "no gaps" algorithm because the key idea is to put the constraint that there is no gap in the value of s_i. Using similar ideas, Neal (2002) extended Gibbs sampling for $s_i' s$ by using a set of auxiliary parameters. Neal's method increases the efficiency in setting a new value of s_i. For a good review of current nonparametric Bayesian models including Dirichlet process in density estimation, regression, survival analysis, hierarchical models and model validation, we refer to Muller and Quintana (2004). For a good review of Markov chain methods for sampling from the posterior distribution of a Dirichlet process mixture model, we refer to Neal (2002).

3.3 Methods

3.3.1 Individual and population activation map

Suppose there are c_j number of individual activation centers for subject j, $j = 1, \ldots, J$. Let $\boldsymbol{\eta}_{jl}^{\mathrm{T}} = (\eta_{1jl}, \eta_{2jl})$ be the l^{th} $(l = 1, ..., c_j)$ activation center for subject j.

We assume that each $\boldsymbol{\eta}_{jl}$ is normally distributed with mean $\tilde{\boldsymbol{\mu}}_{\boldsymbol{jl}}{}^{\mathrm{T}} = (\tilde{\mu}_{1jl}, \tilde{\mu}_{2jl})$ and variance-covariance matrix $\tilde{\boldsymbol{\Sigma}}_{jl} = \begin{bmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{bmatrix}_{jl}$, that is:

$$\pi(\boldsymbol{\eta}_{jl} \mid \tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}) = \phi_2(\boldsymbol{\eta}_{jl}; \tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}). \tag{3.3}$$

where $\phi_2(\cdot; \mathbf{a}, \mathbf{B})$ denotes the density of a bivariate normal distribution with mean $\mathbf{a}$ and covariance matrix $\mathbf{B}$. The normal parameters $\{\tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}\}$ have a Dirichlet process prior of the following form:

$$\{\tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}\} \sim G, \quad G \sim DP(\alpha_0 G_0) \tag{3.4}$$

In this setting, $\{\tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}\}$ are not directly from G_0 but from a random probability distribution G, generated by a Dirichlet process with mean distribution G_0. A key feature of the Dirichlet process is that by marginalizing over G, we can write the prior distribution of the $\{\tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}\}$ in terms of successive conditional distributions. To write the distribution in a convenient way we order $\{\tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}\}$ such that $\{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}\} = \{(\tilde{\boldsymbol{\mu}}_{11}, \tilde{\boldsymbol{\Sigma}}_{11}), \ldots, (\tilde{\boldsymbol{\mu}}_{1c_1}, \tilde{\boldsymbol{\Sigma}}_{1c_1}), \ldots, (\tilde{\boldsymbol{\mu}}_{J1}, \tilde{\boldsymbol{\Sigma}}_{J1}), \ldots, (\tilde{\boldsymbol{\mu}}_{Jc_J}, \tilde{\boldsymbol{\Sigma}}_{Jc_J})\}$, and relabel it using a single index h, $\{\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}}\} = \left\{\left(\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\Sigma}}_1\right), \ldots, \left(\tilde{\boldsymbol{\mu}}_h, \tilde{\boldsymbol{\Sigma}}_h\right), \ldots, \left(\tilde{\boldsymbol{\mu}}_H, \tilde{\boldsymbol{\Sigma}}_H\right)\right\}$. Mathematically, $h = \sum_{k=1}^{j-1} c_k + l$ and $H = \sum_{j=1}^{J} c_j$. With the new notation, the successive conditional distributions are of the following form (Blackwell and MacQueen, 1973):

(3.5)

$$(\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\Sigma}}_1) \sim G_0$$

$$(\tilde{\boldsymbol{\mu}}_H, \tilde{\boldsymbol{\Sigma}}_H) \mid (\tilde{\boldsymbol{\mu}}_1, \tilde{\boldsymbol{\Sigma}}_1), \ldots, (\tilde{\boldsymbol{\mu}}_{H-1}, \tilde{\boldsymbol{\Sigma}}_{H-1}) \begin{cases} = (\tilde{\boldsymbol{\mu}}_h, \tilde{\boldsymbol{\Sigma}}_h) & \text{with probability } \frac{1}{\alpha_0 + H - 1} \\ \sim G_0 & \text{with probability } \frac{\alpha_0}{\alpha_0 + H - 1} \end{cases}$$

Due to the discreteness of G, some of the $\{\tilde{\boldsymbol{\mu}}_h, \tilde{\boldsymbol{\Sigma}}_h\}$ are identical and the H values reduce to some $c_p < H$ distinct values. Write the set of c_p distinct values as $\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\} = \{(\mu_1, \Sigma_1), \ldots, (\mu_{c_p}, \Sigma_{c_p})\}$. These c_p distinct values $\{\boldsymbol{\mu}_i\}$ have the interpretation of being the location of the population activation centers where c_p corresponds to the number of population activation centers. Note that c_p is not fixed but will be determined by the data through the Dirichlet process. We introduce a class variable z_{jl} for each individual activation center, where $z_{jl} = i$ if $(\tilde{\boldsymbol{\mu}}_{jl}, \tilde{\boldsymbol{\Sigma}}_{jl}) = (\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. Now each $\boldsymbol{\eta}_{jl}$ is associated with a particular population center, $\boldsymbol{\mu}_i$, $i = 1, \ldots, c_p$. Given the population activation map $\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$ and conditional on $z_{jl} = i$, $\boldsymbol{\eta}_{jl}$ is bivariate normal with mean $\boldsymbol{\mu}_i$ and covariance matrix $\boldsymbol{\Sigma}_i$. Use of the class parameter z_{jl} greatly increases the computational efficiency of the posterior sampling. We now specify G_0 and α_0 to complete the Dirichlet process prior. Let A_j denote the cross-sectional area of the given MRI slice of the brain of subject j, $j = 1, \ldots, J$. Set $A = \cup_{j=1}^{J} A_j$, the union of individual brain areas. As a best guess of G, $G_0(\boldsymbol{\mu})$ is assumed to follow a uniform distribution over A and $G_0(\boldsymbol{\Sigma})$ takes a conjugate inverse Wishart distribution with 5 degrees of freedom: $\boldsymbol{\Sigma}_i \sim IW_5(\boldsymbol{S}^{-1})$. The precision parameter α_0 of the Dirichlet process has an impact on the posterior distribution of c_p. For a larger α_0, the model puts more weights on G_0 and favors a larger c_p. Following Escobar and West (1995), we give a gamma prior of α_0: $\alpha_0 \sim G(1, 1)$ and update α_0 through the MCMC sampling scheme. In the next section we will describe the distribution of the intensity data given the individual and population activation maps.

3.3.2 Data distribution

Let y_{jv} be the observed effect magnitude at pixel v, $v = 1, \ldots, V$, for subject j, $j = 1, \ldots, J$. We use a single index v to reference a 2D pixel, $\mathbf{x}_v = (x_{1v}, x_{2v})$. Given the individual activation centers $\{\boldsymbol{\eta}_{jl}\}$, we assume that observations are distributed

as a mixture of Gaussian components:

$$f(y_{jv} \mid p_{jvl}, \theta_{jl}, \sigma^2_{jl}, c_j) = p_{jv0}\, \phi(y_{jv}; \theta_0, \sigma^2_0) + \sum_{l=1}^{c_j} p_{jvl}\, \phi(y_{jv}; \theta_{jl}, \sigma^2_{jl}). \tag{3.6}$$

Here, c_j is the number of mixture components for subject j (not including the background component) and $\phi(a; b, c^2)$ is the density at a of a normal distribution with mean b and variance c^2. θ_0 and σ^2_0 represent the mean and variance of the background component. By introducing a latent variable ω_{jv} with $\Pr(\omega_{jv} = l) = p_{jvl}$, the likelihood can be written conditionally as

$$\begin{aligned} f(y_{jv} \mid \theta_0, \sigma^2_0, \omega_{jv} = 0) &= \phi(y_{jv}; \theta_0, \sigma^2_0) \quad \text{for background,} \\ f(y_{jv} \mid \theta_{jl}, \sigma^2_{jl}, \omega_{jv} = l) &= \phi(y_{jv}; \theta_{jl}, \sigma^2_{jl}) \quad \text{otherwise.} \end{aligned}$$

The latent variable ω_{jv} can be interpreted as an allocation variable in the sense that observation y_{jv} is assigned to one of the mixture components with probability p_{jvl}. A pixel is considered active if that pixel belongs to a component with index $l > 0$ and is inactive if that pixel belongs to the background component ($l = 0$).

The mixing weights, p_{jvl}, take the form

$$p_{jvl} \propto \begin{cases} m & l = 0 \\ \phi_2(\mathbf{x}_v; \boldsymbol{\eta}_{jl}, \mathbf{R}_{jl}) & l = 1, \ldots, c_j \end{cases}$$

with $\sum_{l=0}^{c_j} p_{jvl} = 1$. Here $\mathbf{x}_v$ is the spatial location of pixel v and $\mathbf{R}_{jl} = r^2_{jl}\mathbf{I}_2$ where $\mathbf{I}_2$ is the 2×2 identity matrix. The spatial dependence of the data is captured by the weights, p_{jvl}. Given m, the weights largely depend on the distance from the pixel to each of the component centers. This implies that observations that are spatially close to one another are more likely to have similar weights. Hence, spatially close observations will be more correlated than distant observations. Note that an observation distant from all components centers, a priori, will have probability that

it belong to the background center close to 1. Furthermore, if a pixel and component center are coincident, then the a priori probability that this pixel belongs to the background is approximately $m/(r_{jl}^{-2}/2\pi + m)$. fMRI experiments are usually designed such that only a small percentage of the brain (roughly 1%–5%) actually shows any activity. Thus, a priori, we set $m = 19$ and $E(r_{jl}^{-2}) = (2\pi)$. This gives $m/(E(r_{jl}^{-2})/2\pi + m) = 0.95$, reflecting this a priori belief.

The number of mixing components, c_j, for each subject is not known and we estimate it along with all other parameters. A priori we assume c_j is a Poisson random variate with mean 5, for $j = 1, \ldots, J$.

3.3.3 Priors

For mathematical convenience we place inverse gamma (parameterized such that if $x \sim IG(\alpha, \beta)$, $E(x) = \beta/(\alpha - 1)$) priors on r^2, σ^2, σ_0^2 and a Wishart distribution on $\boldsymbol{S}$:

$$r_{jl}^2 \sim IG(2\pi, \beta_r),\ \sigma_{jl}^2 \sim IG(3, \beta_\sigma),\ \sigma_0^2 \sim IG(.001, .001),\ \boldsymbol{S} \sim W_5(\begin{bmatrix} 12 & 0 \\ 0 & 16 \end{bmatrix})$$

Hyperprior distributions are then placed on β_r and β_σ:

$$\beta_r \sim Ga(2, 1), \quad \beta_\sigma \sim Ga(0.01, 0.01)$$

where $Ga(\kappa, \lambda) = \frac{\lambda^\kappa}{\Gamma(\kappa)} x^{\kappa-1} e^{-\lambda x}$ is the gamma distribution. We place rather vague hyperprior distribution on the scale parameters β_σ to reduce the influence of the prior on the posterior estimates of σ_{jl}^2. The Wishart distribution of $\boldsymbol{S}$ gives $E(\Sigma) = \begin{bmatrix} 30 & 0 \\ 0 & 40 \end{bmatrix}$. This implies that, a priori, the spread of the individual component centers in the x direction is about $5.5(= \sqrt{30})$ in the x direction and about $6.3(= \sqrt{40})$ in the y direction. Given the brain dimension of 53×63, the spread is about 1/10 of the range.

It remains to specify the priors of the intensity parameters, θ_{jl}, in the likelihood specified in equation (2.1). Recall that θ_{jl}, $l > 0$, are the mean intensity levels of the individual mixture components for active pixels and θ_0 is the mean intensity level of inactivated pixels. We choose the following priors:

$$\theta_0 \sim N(0,1) \quad \text{and} \quad \theta_{jl} \sim \text{trunc}_{(0,\infty)} N(\lambda_\theta, \sigma_\theta^2),$$

where

$$\lambda_\theta \sim N(35, 1 \times 10^8) \quad \text{and} \quad \sigma_\theta^2 \sim IG(.01, .01).$$

Furthermore, $\text{trunc}_{(0,\infty)} N(\lambda_\theta, \sigma_\theta^2)$ denotes the normal distribution with mean λ_θ and variance σ_θ^2 truncated to $(0, \infty)$.

3.4 Posterior estimation

To sample from the posterior distribution we propose the following Markov chain Monte Carlo (MCMC) steps: (1) add or remove an individual component (2) update the class variable z for each individual component (3) change $(\boldsymbol{\eta}, r^2, \theta, \sigma^2)$ for an individual component (4) change $(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ for a population center (5) update the class variable ω for each observation, and (6) update hyperparameters $(\beta_r, \beta_\sigma, \lambda_\theta, \sigma_\theta^2, \boldsymbol{S}, \alpha_0)$. We will briefly describe each step in this section. Details of the algorithm are included in the Appendix. In Step (1) we propose to add a new individual component with probability 0.5 and propose to delete an individual component with probability 0.5 at each iteration. When we propose to add an individual component the parameters defining the new component are drawn from their prior distributions. This involves drawing $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ from expression (3.5). The use of the prior distributions in proposing new values leads to a simplification of the acceptance probability. In Step (2) we use Algorithm 8 proposed by Neal (2002). The class variable z is sampled using a Gibbs type sampler. Note that both step (1) and step (2) can change the number of

population centers if either a new value of $(\tilde{\boldsymbol{\mu}}, \tilde{\boldsymbol{\Sigma}})$ is created or an existing population center becomes empty. Step (3)-(6) do not involve any change in the dimension of the parameter space. Conditional on the number of population centers and individual components we update $(\boldsymbol{\eta}, r^2, \boldsymbol{\mu})$ using a random walk Metropolis-Hastings (MH) step. The variances in the proposal distribution for the MH steps were chosen to obtain acceptance rates of approximately 35%. $(\theta, \boldsymbol{\Sigma}, \sigma^2, \beta_r, \beta_\sigma, \lambda_\theta, \sigma_\theta^2, \boldsymbol{S}, \alpha_0)$ is sampled using a Gibbs step. According to Escobar and West (1995) the Gibbs step for sampling α_0 has two parts. First sample a beta variable β_0 given current value of α_0 and then sample a new α_0 from a mixture of two gamma densities. Following Fernandez and Green (2002) we use the marginal expression for the likelihood as specified in equations (2.1) to obtain better mixing.

3.5 Results

We apply the infinite mixture model to the working memory data described at the beginning of the thesis. Figure 3.1 and 3.2 show the results. The posterior probability of activation obtained from this model is very similar to that obtained from the finite mixture model. Signals are found for subject 1, 3, and 9 but not found for subject 20. However, the infinite mixture model fits less individual component centers for each activation area. For the population level results, the 2D posterior rate image of the population center location (Figure 3.2 (a)) shows 6 population centers. Each center contains 2 to 3 pixels. The Population Center Prevalence image in Figure 3.2 (b) shows that 3 of these 6 population centers are present in more than 50% of the subjects. The other 3 population centers are present in 20-30% of the subjects. Based on the Population Center Area image (Figure 3.2 (h)), the middle and the upper 2 population centers have larger average areas, which is about

10-15 pixels. The bottom three population centers have smaller average areas, which contain about 5 pixels. The infinite mixture model finds similar population center locations as does the finite mixture model. But the Individual Component Posterior image is different. The infinite mixture model shows that individual component centers are more spread in the y direction. The Population Center x-Scale and the Population Center y-Scale images show the spread of the individual component center in the x and the y direction. For most pixels, the spread in x direction (σ_{11}) is 1-2 and is 2-3 in the y direction (σ_{22}). The trace plots of the number of population centers show that the infinite mixture model converges faster and mix better than the finite mixture model. Results for all 21 subjects are given in Appendix D.

The main contributions of this work are as follows. A traditional voxel-based approach can only detect activated voxels for which the average signal is significant across subjects. However, subjects sometimes don't have, or only a small percentage of the total subjects have, overlapped activated voxels even though these activated voxels are close to each other and represent the same anatomical/functional area. Our models fit individual component centers for subject level activation areas and find population activation centers based on these individual component centers. We explicitly model the location of the population activation center with a separate parameter describing the variability of the individual component centers. Unlike the voxel-based approach our methods allow for variation in both signal intensity and activation location at the individual level. The location of the population activation centers are summarized using the posterior rate function. This provides better precision in the estimation of the population center location than the standard mass-univariate method.

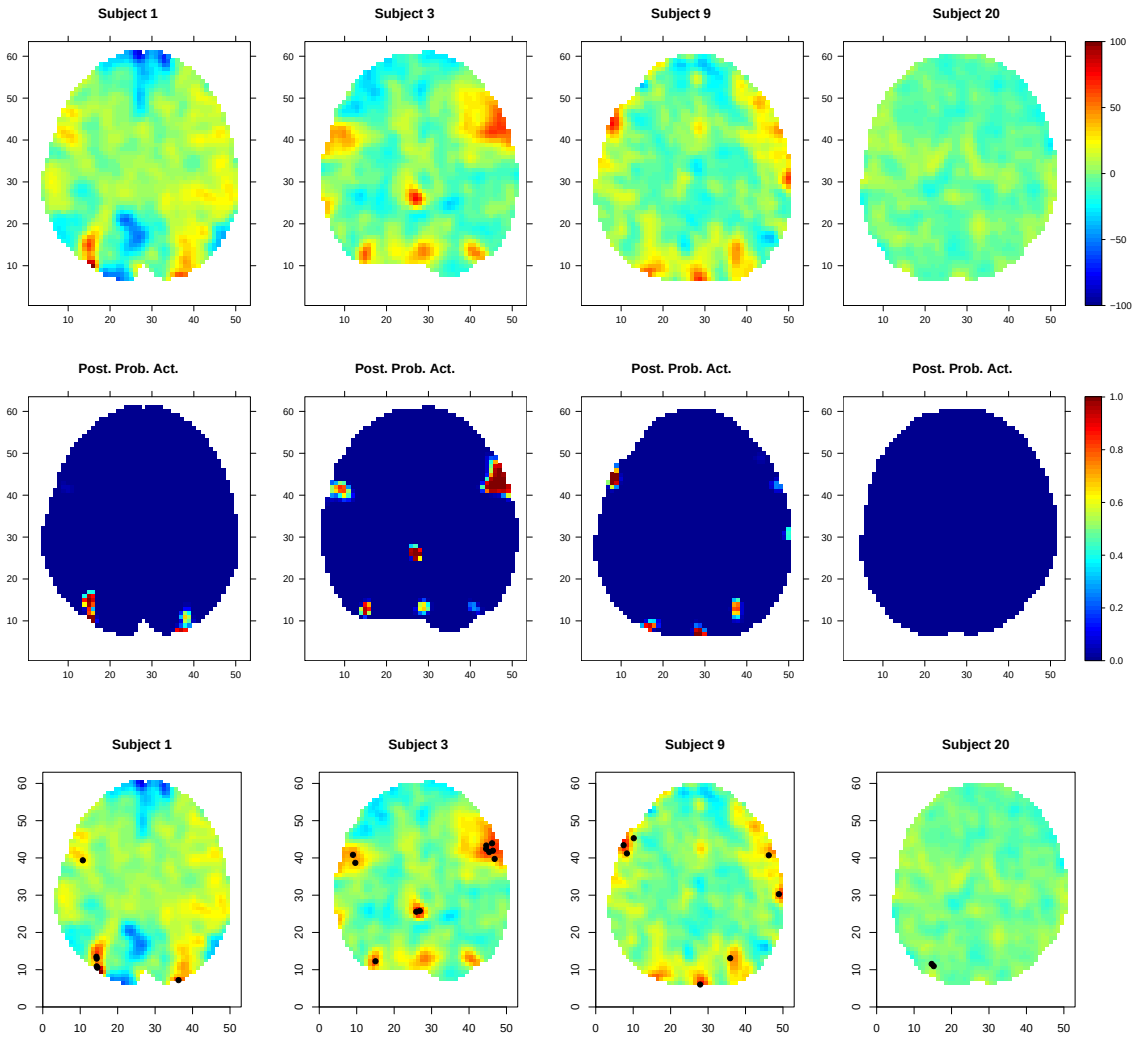


Figure 3.1: Top row: The intensity data from slice 21 for 4 subjects. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

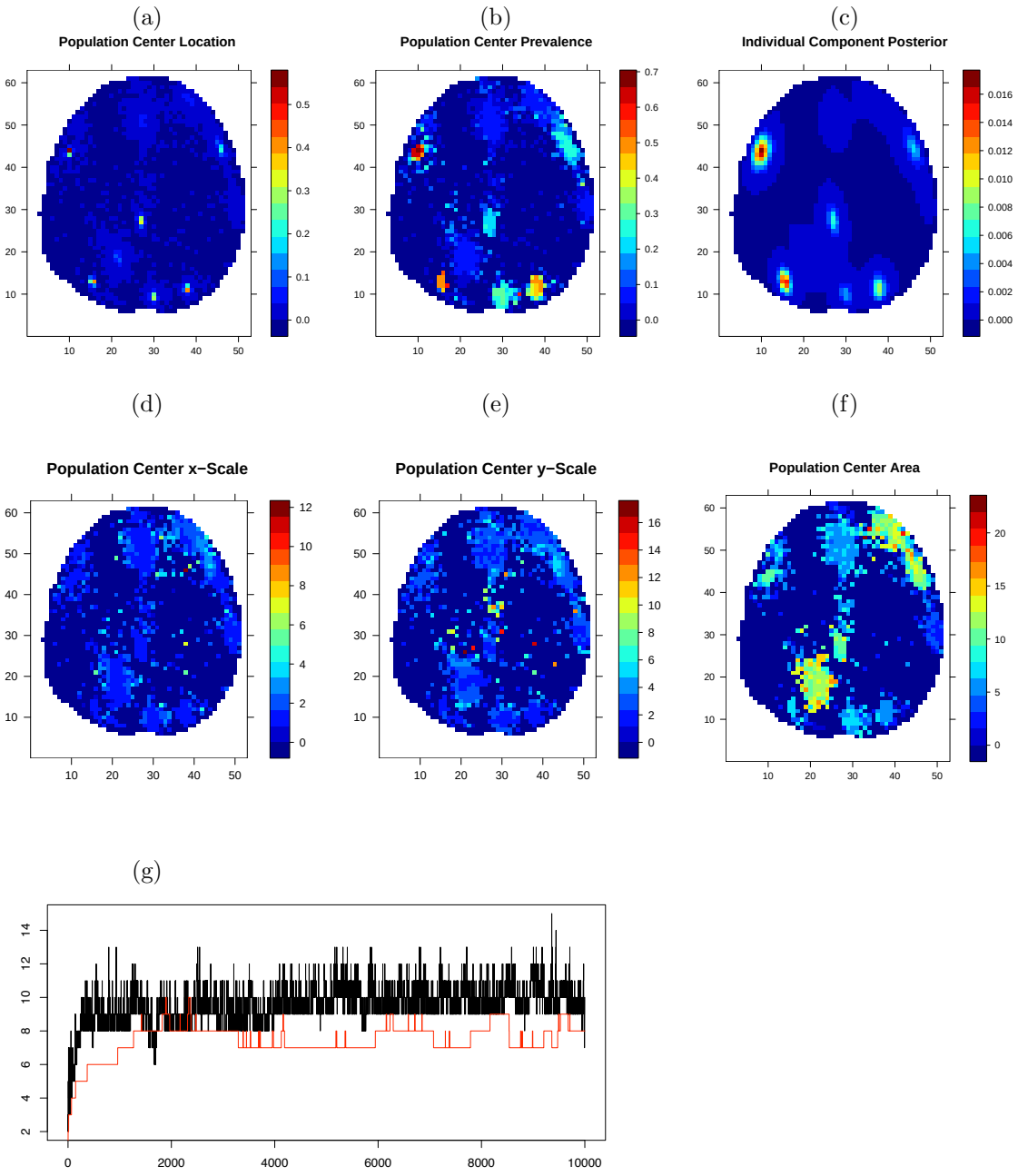


Figure 3.2: (a) 2D posterior rate function of population center locations; (b) proportion of subjects containing evidence for a population center; (c) posterior-predictive density of individual center locations $p(\tilde{\boldsymbol{\eta}} \mid \mathbf{y})$; (d) standard deviation of individual components about population centers on the x direction; (e) standard deviation of individual components about population centers on the y direction; (f) the average number of activated pixels for a population center; (g) trace plot of the number of population centers for the infinite mixture model (black) and the finite mixture model (red), starting from 1 center.

CHAPTER IV

SIMULATION STUDIES AND SENSITIVITY ANALYSIS

In the previous two chapters we introduce two Bayesian models. Both are RJMCMC Bayesian hierarchical models (In Chapter II the finite mixture model and in Chapter III the infinite mixture model with the Dirichlet process prior). To investigate the performance of our Bayesian models we perform several simulation studies. For each simulated data set we apply both models and compare the results. We also carry out sensitivity analyses to investigate how hyperparameter values influence the posterior distributions. We present simulation results in Section 4.1, while results from the sensitivity analysis are given in Section 4.2.

4.1 Simulation Studies

4.1.1 Noise data

In this section, we simulate a data set which consists of 10 simulated images with no signal, each under the following three conditions:

(a) all images contain independent N(0,1) noise.

(b) smooth the independent image with a symmetric Gaussian kernel with standard deviation 0.5.

(c) smooth the independent image with a symmetric Gaussian kernel with standard deviation 1.5.

Condition (b) and (c) create images with correlated intensities. As an example, Figure 4.1 (top row) shows one image under each condition. We run both the finite mixture model and the infinite mixture model for each of 20 simulated data sets (each data set consists of 10 images). Figure 4.1 (bottom row) shows the mean image of all 20 data sets. We summarize results by taking the average over all 20 data sets. We find no activation at either the individual level or the population level. The posterior w is zero for most subjects and close to zero for the rest. The number of population and the number of individual component centers are not zero. But the addition and deletion of the centers follow a random pattern over the entire image. This can be seen from Figure 4.2 which displays the posterior distribution of the location of the population centers and the posterior predictive density of the location of the individual component center.

4.1.2 Simulated multiple activation centers with spherical Gaussian for η

Each simulated data set consists of 10 images, each with multiple activation centers added to the same three conditions as in last section:

(a) all images contain independent intensities.

(b) smooth the independent image with a symmetric Gaussian kernel with standard deviation 0.5.

(c) smooth the independent image with a symmetric Gaussian kernel with standard deviation 1.5.

The parameters in the simulation include different numbers of individual activation components and include centers with different mean intensity levels and different sizes. There are three population activation centers located roughly in the upper left, upper right and lower center portion of the image. For each of the population

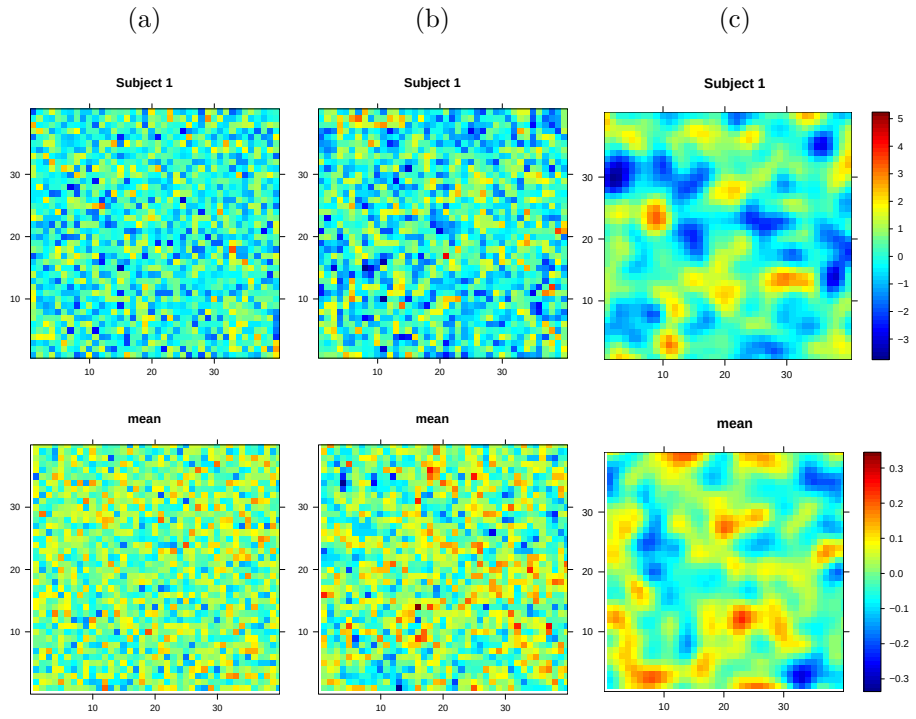


Figure 4.1: Simulated Noise Data under three conditions: (a) independent noise; (b) smoothed with Gaussian standard deviation = 0.5; (c) smoothed with Gaussian standard deviation = 1.5. First row: The simulated data for one instance. Second row: Mean image averaging over all 20 runs.

centers, we randomly put 10, 5 and 1 individual component centers associated with the upper left, upper right and the lower center population centers, respectively. The location of the individual component center $\boldsymbol{\eta}$ is drawn from a spherical Gaussian, i.e. $\boldsymbol{\eta}_j \mid z_j = i, \boldsymbol{\mu}_i, \tau_i^2 \sim N(\boldsymbol{\mu}_i, \tau_i^2 \mathbf{I}_2)$. Inactivate pixels are drawn from a $N(0,1)$ distribution. Parameter values are listed in Table 4.1. Figure 4.3 (top row) shows an image under each condition. We run both the finite mixture model and the infinite mixture model for each data set and repeat the simulation 20 times. Figure 4.3 (bottom row) shows the mean image of the simulated data after averaging over all 20 data sets. We summarize results in Figure 4.4 by taking an average over all 20 simulations. Results are similar for both the finite mixture model and the infinite

i	1	2	3
μ_i	$\begin{pmatrix} 10 \\ 30 \end{pmatrix}$	$\begin{pmatrix} 32 \\ 25 \end{pmatrix}$	$\begin{pmatrix} 15 \\ 10 \end{pmatrix}$
τ_i^2	4	4	1
z_{jl}	1	2	3
θ_{jl}	5	10	10
r_{jl}^2	1.5	1	0.5
σ_{jl}^2	1	4	4

Table 4.1: Parameters used to simulate the multiple activation data set.

mixture model. In the Individual Component Posterior image, the locations of the three major modes are close to the true population values. Note that there are 3 modes, one for each population center. The largest mode occurs in the upper left where there are 10 individual component centers. The Population Center Location image gives similar information about the location of the 3 population centers as does the Individual Component Posterior image. However, note that the variability of the population locations is much smaller than the variability across subjects. In fact, the population level information is quite precise.

To compare our methods with the classical method, we take the Population Center Location image and calculate the root mean square error (RMSE) for each population activation center. The RMSE is defined as the distance from the estimated population center location to the true population center location averaging over all 20 data sets. Let $\boldsymbol{\mu}_i$, $i = 1, 2, 3$ be the true location of the three population centers. In this simulation, $\boldsymbol{\mu}_i = (15, 10)$, the bottom center, $\boldsymbol{\mu}_2 = (32, 25)$, the upper right center, and $\boldsymbol{\mu}_3 = (10, 30)$ the upper left center. The estimated population center location $\hat{\boldsymbol{\mu}}_i$, $i = 1, 2, 3$ is derived by taking the spatial location of the local maximum of each of the three activation areas using the Population Center Location images.

		Population Center Location		
		Bottom	Upper right	Upper left
Classical t-image	unsmoothed	9.39	10.91	4.27
	smoothed with SD=0.5	7.98	10.66	3.11
	smoothed with SD=1.5	9.88	9.23	1.96
Finite mixture model	unsmoothed	4.17	4.47	3.63
	smoothed with SD=0.5	2.55	1.55	1.32
	smoothed with SD=1.5	1.92	1.55	1.30
Infinite mixture model	unsmoothed	2.61	1.50	1.43
	smoothed with SD=0.5	1.96	1.72	1.40
	smoothed with SD=1.5	1.84	1.55	1.23

Table 4.2: The root mean square error (RMSE) from the classical t-image, the finite mixture model and the infinite mixture model.

The $RMSE_i$, $i = 1, 2, 3$ is of the following form:

$$RMSE_i = \sqrt{\sum_{k=1}^{20} (\hat{\boldsymbol{\mu}}_{ik} - \boldsymbol{\mu}_{ik})^T (\hat{\boldsymbol{\mu}}_{ik} - \boldsymbol{\mu}_{ik})/20} \tag{4.1}$$

We compare the RMSE for the finite mixture and the infinite mixture model using the Population Center Location images with the RMSE using the classical t-image. Results are summarized in Table 4.2. For both unsmoothed and smoothed data sets the finite mixture model and the infinite mixture model perform dramatically better than the classical t-image in terms of the precision of the estimated population center locations for all three activation areas. The t-image gives a reasonably small RMSE only for the upper left population center and under the condition when the data sets are smoothed with Gaussian deviation of 1.5. The upper left population center is the center that has the most individual components. So the upper left center has the most support. The smoothness makes the activation even stronger. In fact within each model the RMSE is the smallest for this category. Another finding is that for the unsmoothed data sets, the infinite mixture model still gives good estimation of the population center location, better than the finite mixture model.

4.1.3 Simulated multiple activation centers with non-spherical Gaussian for η

In this section we compare the performance of the finite mixture model with the infinite mixture model on a simulated data set with a non-spherical Gaussian for the location of the individual component centers. We simulate data set which consists 16 images with multiple activation centers. The top row of Figure 4.5 shows four of the simulated data sets. There are three population activation centers. images 1 through 9 possess activations corresponding to all three centers. images 10 through 12 possess activations corresponding to only two of the three population level centers. images 13 through 15 possess only one activation region each. image 16 has no activation. The actual locations of the population centers and the individual components are summarized in Figure 4.7(a). The mean intensities of these three population level centers are 15, 20 and 25. The variance of the relative intensity for activated pixels are, respectively, 1, 4 and 9. We set $\mathbf{\Sigma}_i = \begin{bmatrix} 8 & 5 \\ 5 & 4 \end{bmatrix}, \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}$ and $\begin{bmatrix} 1 & -0.5 \\ -0.5 & 1 \end{bmatrix}$ for $i = 1, 2, 3$ to obtain different spreads of the individual component centers. Σ_1 gives a population center with positively correlated ($\rho \approx 0.884$) individual component centers. The spread in the x-direction is larger than the spread in the y-direction. Σ_2 gives a population center with independent individual component centers with equal spread in the x and y directions. Σ_3 gives a population center with negatively correlated ($\rho = -0.5$) individual component centers with equal spread in the x and y direction. The variance of the background is 4. We set $r_{jl}^2 = 1.5$, 1 and 0.5 for all j for the three different individual activation components $l = 1, 2, 3$ to obtain different sizes for each individual component. We run the chain for 10,000 iterations and discard the first 5,000 as burn-in. We save every 5th iteration for the results. To assess convergence we start at different initial values. There are no significant

differences in the results.

Figure 4.5 and 4.6 display the individual level results for the finite mixture and the infinite mixture model. In both Figures, the second row shows the Posterior Probability of Activation images. The third row shows the posterior distributions of the number of individual component centers. The bottom two rows show the center location of individual components at iteration 7000 and at iteration 7250. Both models give similar results at the individual level. Activated pixels are correctly classified. Both models fit several component centers for one activation area. However, the finite mixture model fits more individual components than the infinite mixture model. By comparing the center location of individual components at iteration 7000 to those at iteration 7250, we can see that individual component centers stay consistently around activated area, but move randomly in background areas. Results for all 16 images can be found in Appendix E and Appendix F.

Figure 4.7 displays the population level results for both models. For comparison we display in Figure 4.7(a2) the classical t-image and in Figure 4.7(a3) the $-\log_{10}$P-value image. Figure 4.7 (b1)(c1) show the 2D posterior rate function of the population center locations. The infinite mixture model gives three major peaks for the location of the population centers, which are about at the true values. The finite mixture model gives similar results for the location of the bottom and the upper right population center, but fits two more population centers for the upper left area. This is due to the positively correlated individual component centers around the population center. The finite mixture model which assumes an independent structure on the covariance of the normal mixture for $\boldsymbol{\eta}$ has to fit more population centers to capture the elliptical shape of that area. Figure 4.7 (b3)(c3) give an example of where the population centers and where the individual component centers are

actually located for each model. Figure 4.7 (b2)(c2) show the posterior predictive density of the individual component center, which give information about how these two models capture the different spread of the individual component centers about the population center. Similar to the real data analysis, the infinite mixture model results in c_p mixing better and converging faster for the simulated data (see Figure 4.7 (b4)(c4) for the trace plots).

4.2 Sensitivity Analysis

4.2.1 Results of sensitivity analysis for the finite mixture model

We study the sensitivity of our approach to the choice of hyperprior specifications. Our main interest is in how the population level results are related to various prior specifications. We study the sensitivity of the posterior to the five following scenarios:

(a) $m = 19$, $c_j \sim P(5)$, $j = 1, \ldots, J$, $\lambda_p = 5A^{-1}$, $\alpha_\tau = \alpha_\sigma = 3$, $\lambda_\theta \sim N(35, 10^8)$, the default setting.

(b) same as (a) but with different random seeds (different initial values).

(c) same as (a) but with $m = 5\frac{2}{3}$, which corresponds to an a priori probability that a pixel belongs to the background center of approximately 0.85.

(d) same as (a) but with $\lambda_p = 10A^{-1}$ and $c_j \sim P(10)$, $j = 1, \ldots, J$.

(e) same as (a) but with $\alpha_\tau = \alpha_\sigma = 2$, $\lambda_\theta \sim N(70, 10^8)$.

Overall we find no appreciable differences in the population level results (i.e. Figures 4.8 (a) through (e)), under all five scenarios. As expected, scenarios (a) and (b) give almost identical results for both the individual and the population level. For scenario (c), as m decreases the probability that a pixel belongs to the background decreases. This encourages more individual components at low intensity areas. Also, the posterior distribution of the population centers is slightly influenced. We observe

a shift in the posterior distribution of the number of individual components, c_j, $j = 1, \ldots, J$, and in the distribution of the number of population centers c_p. The increase of k in scenario (d) brings the same effect to the posterior distribution of c_p and c_j. This influence of the Poisson mean on the posterior distribution of the number of components is expected, see, e.g., Green (1995).

4.2.2 Results of sensitivity analysis for the infinite mixture model

In the infinite mixture model, we no longer specify a Poisson prior for the number of population centers, c_p. Instead, the prior believe of c_p is controlled by one of the Dirichlet process parameters, α_0. We make corresponding changes in the sensitivity analysis. We remove the settings of λ_p and add one more scenario for α_0. We also add two more scenarios for $\boldsymbol{S}$, the covariance matrix controlling the spread of the individual component centers. This gives the following eight scenarios:

(a) $m = 19$, $c_j \sim P(5)$, $j = 1, \ldots, J$, $\alpha_\tau = \alpha_\sigma = 3$, $\lambda_\theta \sim N(35, 10^8)$, $\alpha_0 \sim Ga(1, 1)$, $\boldsymbol{S} \sim W_5(\begin{bmatrix} 12 & 0 \\ 0 & 13 \end{bmatrix})$ the default setting.

(b) same as (a) but with different random seeds (different initial values).

(c) same as (a) but with$\boldsymbol{S} \sim W_5(\begin{bmatrix} 24 & 0 \\ 0 & 24 \end{bmatrix})$.

(d) same as (a) but with $\boldsymbol{S} \sim W_5(\begin{bmatrix} 18 & 0 \\ 0 & 6 \end{bmatrix})$.

(e) same as (a) but with $m = 5\frac{2}{3}$, which corresponds to an a priori probability that a pixel belongs to the background center of approximately 0.85.

(f) same as (a) but with $c_j \sim P(10)$, $j = 1, \ldots, J$.

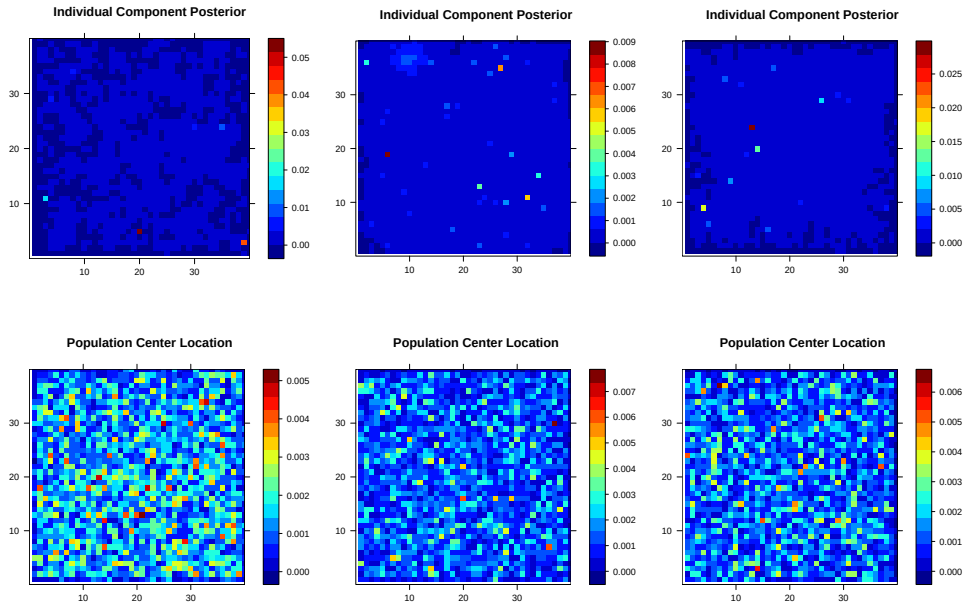
(g) same as (a) but with $\alpha_\tau = \alpha_\sigma = 2$, $\lambda_\theta \sim N(70, 10^8)$.

(h) same as (a) but with $\alpha_0 \sim Ga(4, 2)$.

The results (displayed in Figures 4.9 (a) through (d) and Figures 4.10 (e) through (h)) are similar to these for the finite mixture model. The posterior distribution of the location of the population activations are almost the same under all eight scenarios. The posterior distribution of the number of population and the individual component centers are slightly different under different scenarios. For scenario (e), as m decreases the posterior distribution of the number of individual components, c_j, $j = 1, \ldots, J$, and the number of population centers c_p shift slightly to a higher value. For scenario (h) the mean of α_0 increases from 1 to 2. This favors the addition of a new population center. However, we do not observe an increase in the number of population centers in the posterior distribution.

(a) (b) (c)

Finite Mixture Model



Infinite Mixture Model

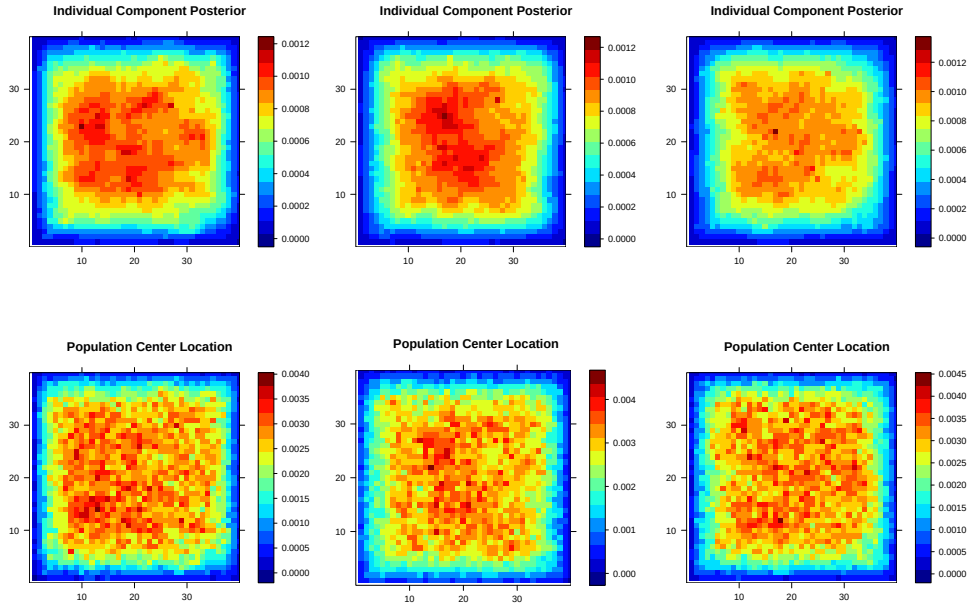


Figure 4.2: Simulation results for the noise data under three conditions: (a) independent noise; (b) smoothed with Gaussian standard deviation = 0.5; (c) smoothed with Gaussian standard deviation = 1.5. First two rows: results from the finite mixture model. The posterior predictive density of $\boldsymbol{\eta}$ and the posterior distribution of $\boldsymbol{\mu}$. Bottom two rows: results from the infinite mixture model.

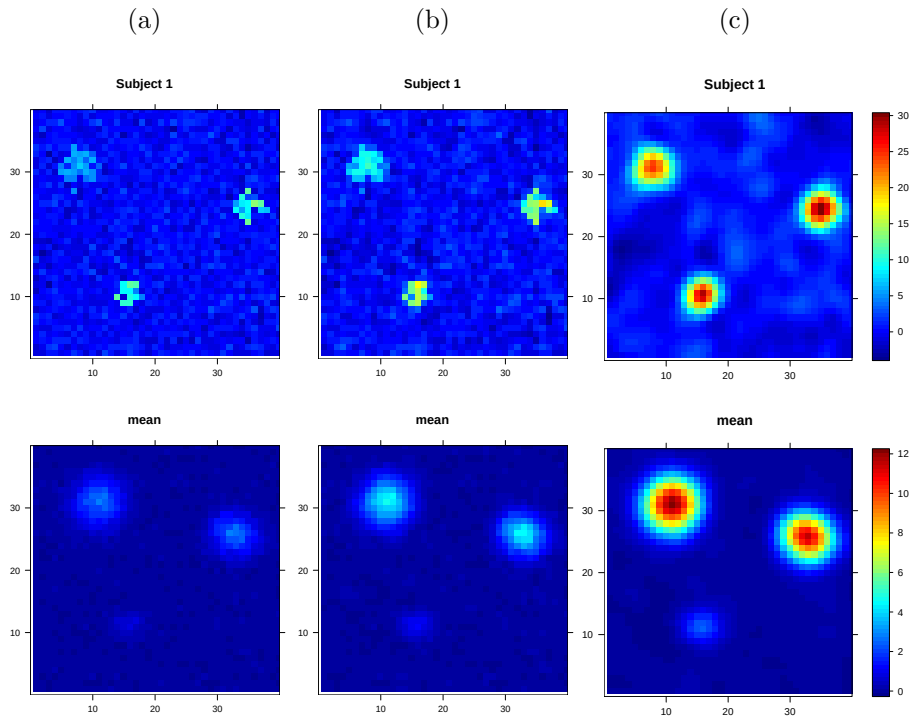
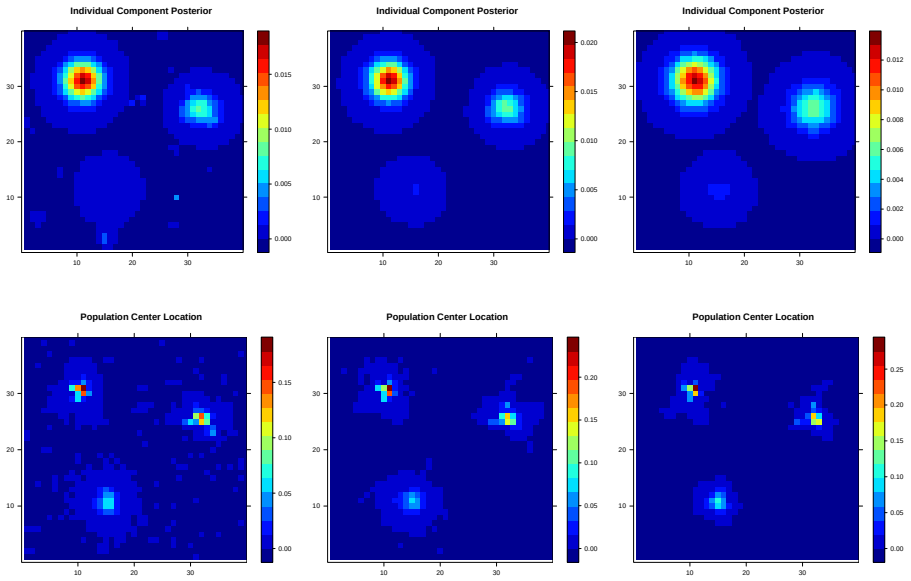


Figure 4.3: Simulated signal data sets with Spherical Gaussian under three conditions: (a) independent signals; (b) smoothed with Gaussian standard deviation = 0.5; (c) smoothed with Gaussian standard deviation = 1.5. First row: The simulated data for image 1. Second row: Mean image averaging over all 20 data sets.

(a) (b) (c)

Finite Mixture Model



Infinite Mixture Model

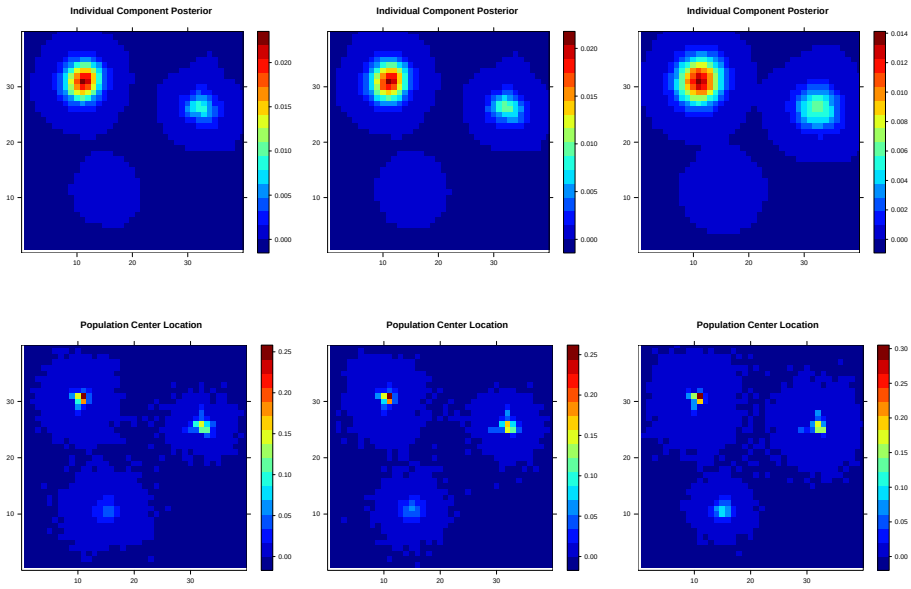


Figure 4.4: Simulation results for the signal data with Spherical Gaussian under three conditions: (a) independent noise; (b) smoothed with Gaussian standard deviation = 0.5; (c) smoothed with Gaussian standard deviation = 1.5. First two rows: results for finite mixture model. The posterior predictive density of $\boldsymbol{\eta}$ and the posterior distribution of $\boldsymbol{\mu}$. Bottom two rows: results for infinite mixture model.

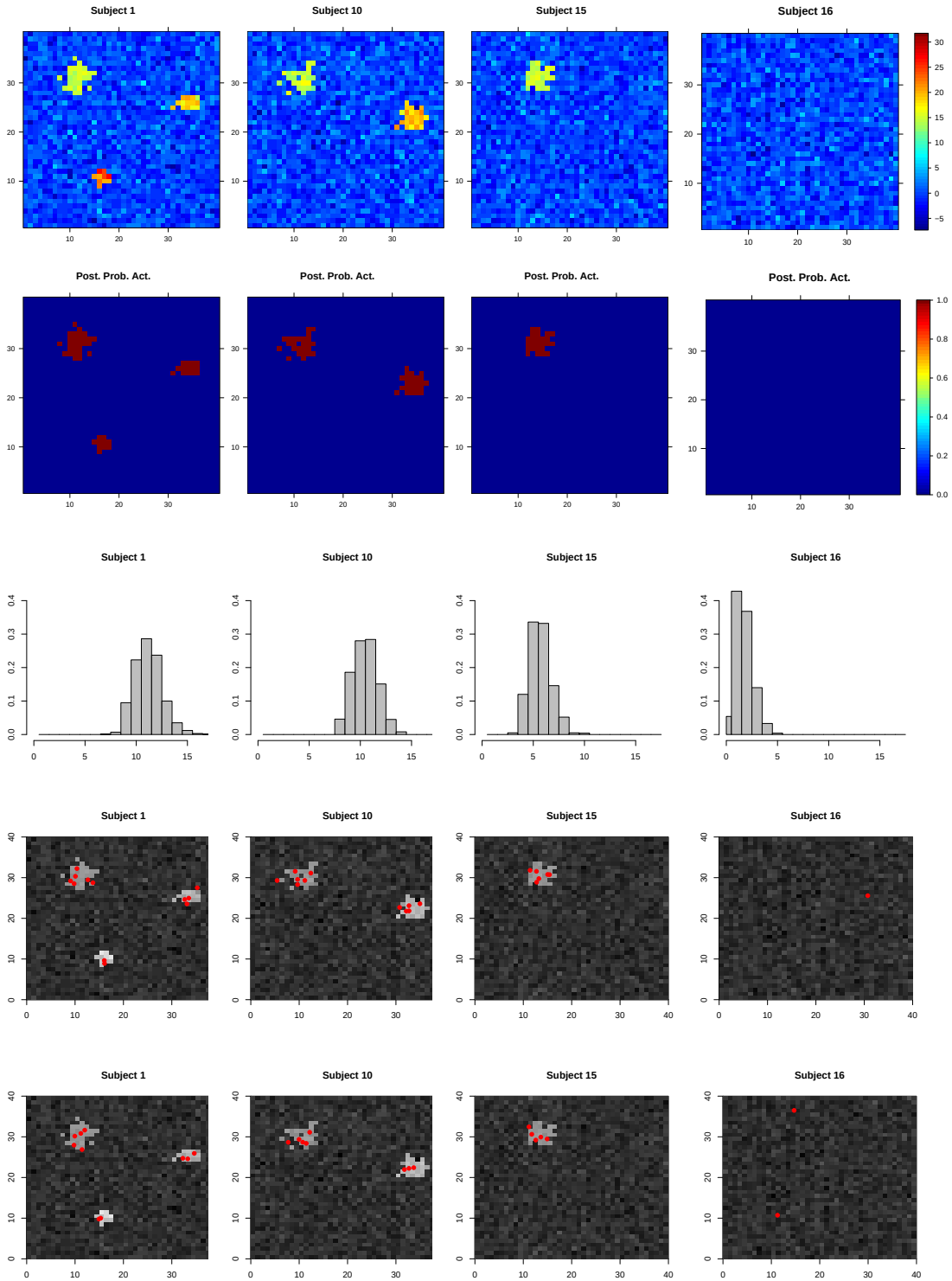


Figure 4.5: Simulation results for the infinite mixture model (using the signal data with non-spherical Gaussian). First row: The intensity data for subject 1, 10, 15 and 16. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Third row: the posterior distributions of the number of individual component center. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

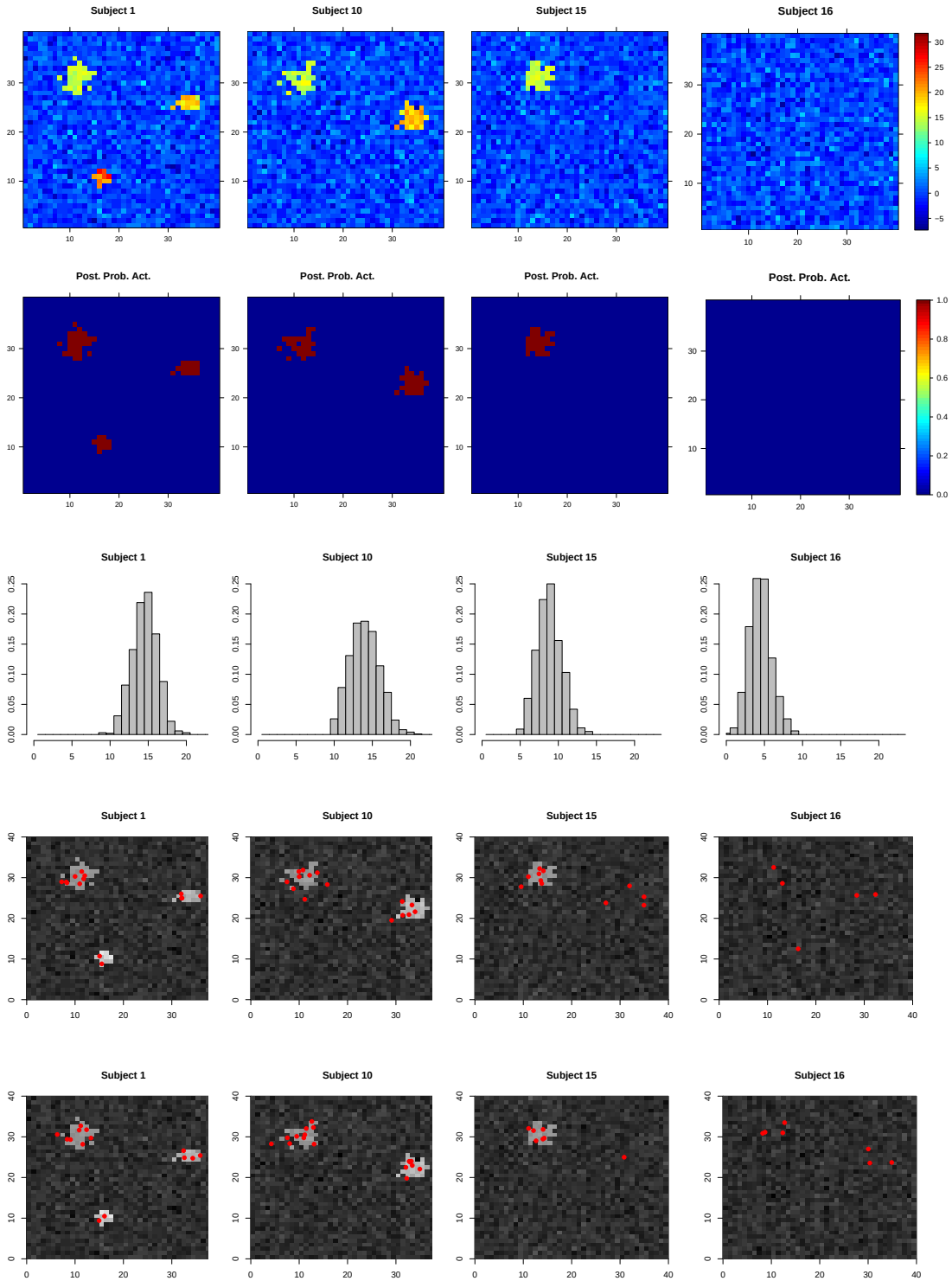


Figure 4.6: Simulation results for the finite mixture model (using the signal data with non-spherical Gaussian). First row: The intensity data for subject 1, 10, 15 and 16. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Third row: the posterior distributions of the number of individual component center. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

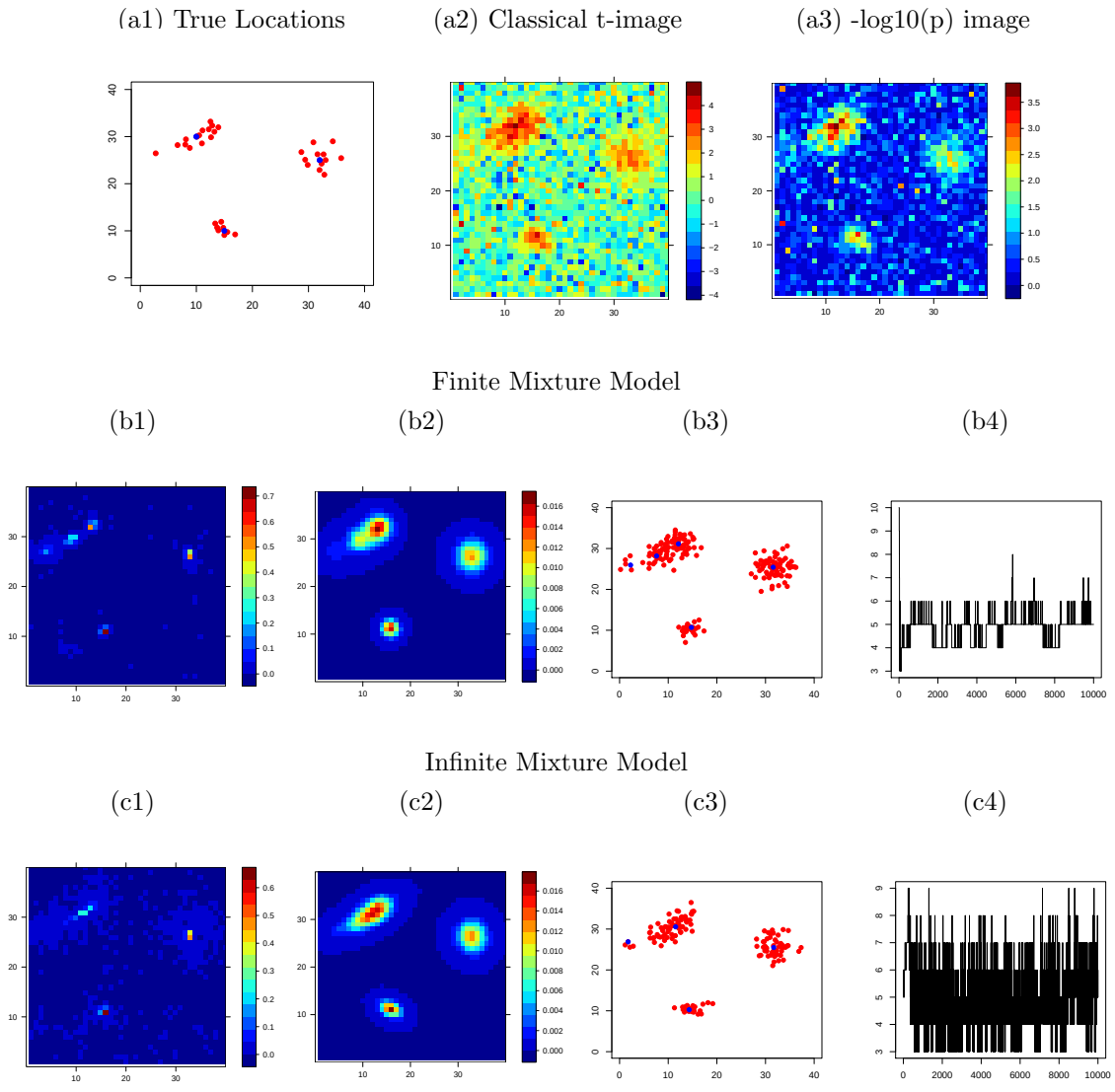


Figure 4.7: Simulation results for the signal data with non-spherical Gaussian: (a1) True locations of population and individual centers for simulation; (a2) Classical one-sample t-test. (a3) minus log base 10 p-values from the t-image; (b1)(c1) 2D posterior rate function of population center locations; (b2)(c2) posterior-predictive density of individual center locations $p(\tilde{\boldsymbol{\eta}} \mid \mathbf{y})$; (b3)(c3) posterior location of population and individual centers at iteration 7000; (b4)(c4) trace plot of the number of population centers c_p

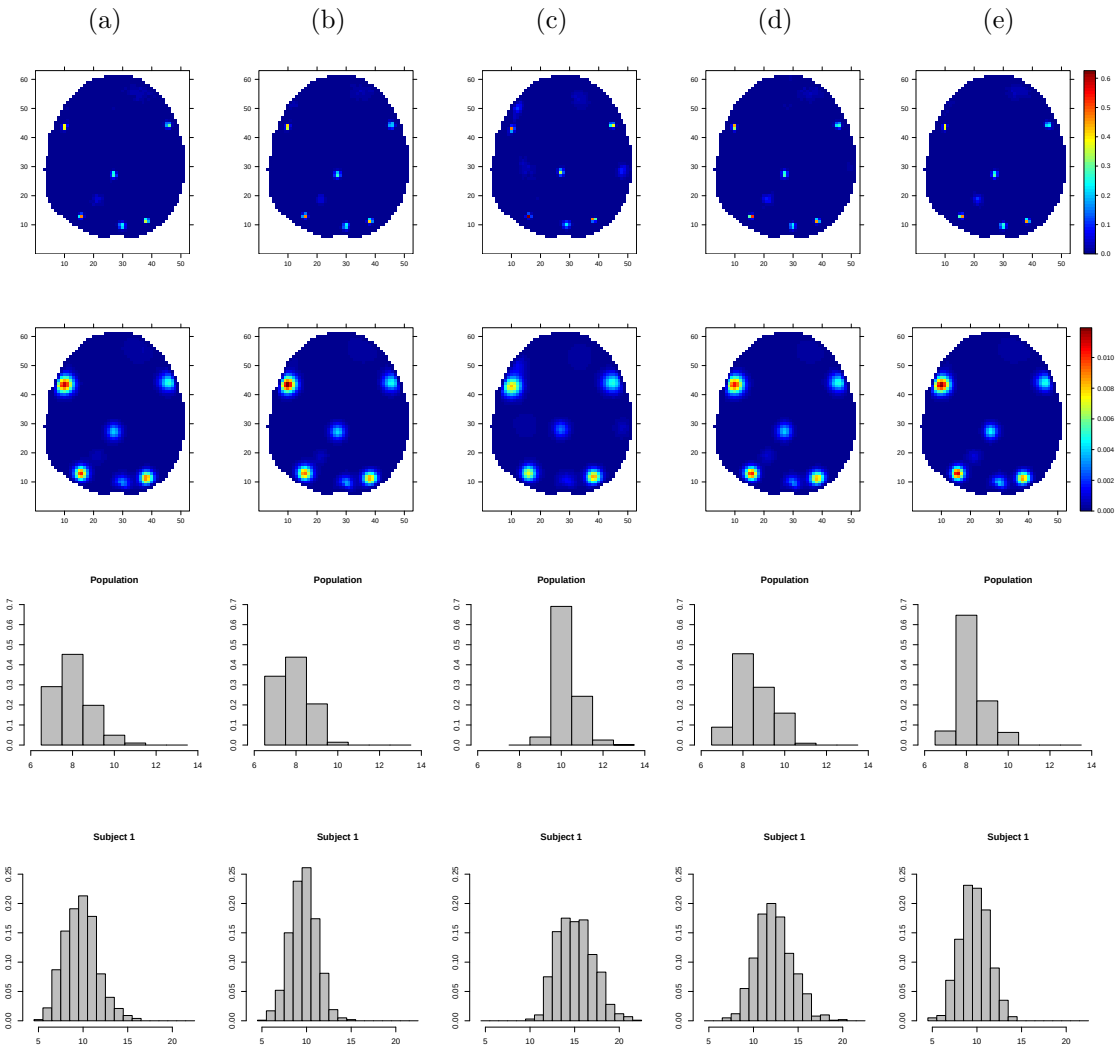


Figure 4.8: Sensitivity analysis results for the finite mixture model under five scenarios. (a)(b) $k = k_p = 5, m = 19$, $\alpha_\tau = \alpha_\sigma = 3$, $\lambda_\theta \sim N(35, 10^8)$; (c) $m = 5\frac{2}{3}$ (d) $k = k_p = 10$; (e) $\alpha_\tau = \alpha_\sigma = 2$, $\lambda_\theta \sim N(70, 10^8)$. First row: 2D posterior rate function of population center locations; Second row: Posterior-predictive density of individual center locations $p(\tilde{\boldsymbol{\eta}} \mid \mathbf{y})$; Bottom two rows: Posterior distributions of the number of population centers and the center of individual components for subject 1

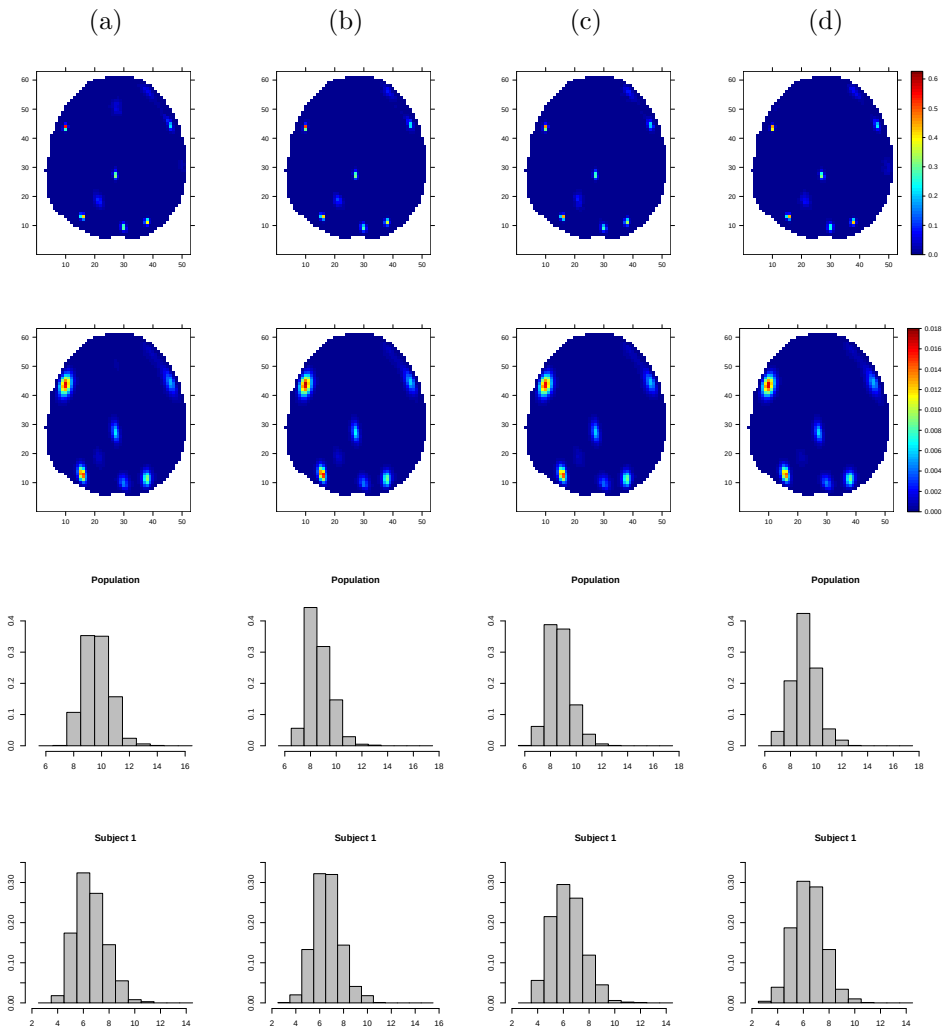


Figure 4.9: Sensitivity analysis results for the infinite mixture model under eight scenarios (to be continued on the next page). (a)(b) $k = 5, m = 19, \alpha_\tau = \alpha_\sigma = 3, \lambda_\theta \sim N(35, 10^8), \alpha_0 \sim Ga(1,1), \boldsymbol{S} \sim W_5(diag(12,13)))$; (c) $\boldsymbol{S} \sim W_5(diag(24,24))$; (d)$\boldsymbol{S} \sim W_5(diag(18,6))$. First row: 2D posterior rate function of population center locations; Second row: Posterior-predictive density of individual center locations $p(\tilde{\boldsymbol{\eta}} \mid \mathbf{y})$; Bottom two rows: Posterior distributions of the number of population centers and the center of individual components for subject 1

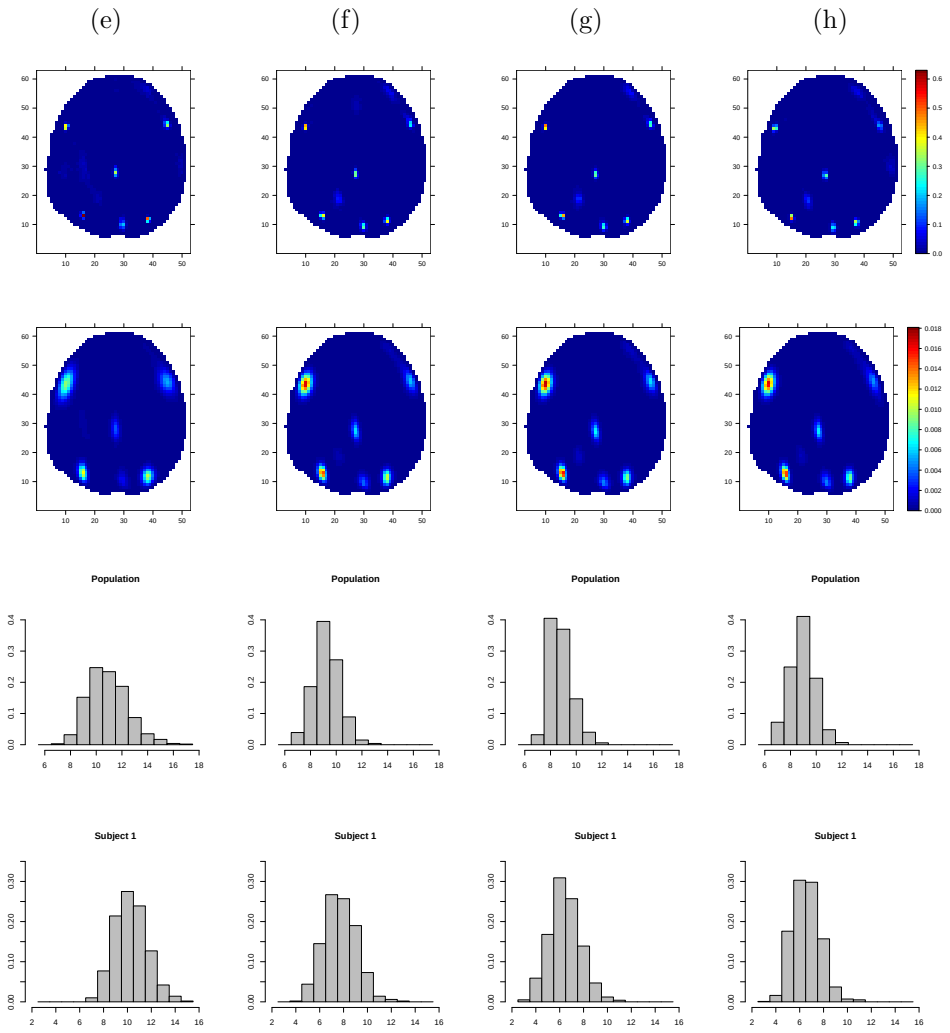


Figure 4.10: Sensitivity analysis results for the infinite mixture model under eight scenarios (continued from the previous page). (e) $m = 5\frac{2}{3}$ (f) $k = 10$; (g) $\alpha_\tau = \alpha_\sigma = 2$, $\lambda_\theta \sim N(70, 10^8)$; (h) $\alpha_0 \sim Ga(4, 2)$. First row: 2D posterior rate function of population center locations; Second row: Posterior-predictive density of individual center locations $p(\tilde{\boldsymbol{\eta}} \mid \mathbf{y})$; Bottom two rows: Posterior distributions of the number of population centers and the center of individual components for subject 1

CHAPTER V

FUTURE WORK AND DISCUSSION: SPATIAL CLUSTER MODELLING BASED ON COX PROCESSES

5.1 Introduction

We have described two Bayesian mixture models for fMRI data analysis: the "finite mixture model" and the "infinite mixture model". Both methods consider the spatial structure of the signal, which is often ignored in frequentist approaches. Moreover this method extends the current fMRI Bayesian spatial modeling literature in two key ways. First, we consider multi-subject data, explicitly model population centers and the dispersion of individual's response about those centers. Second, instead of assuming a normal shape model for activation magnitude, we assume a normal shape model for probability of activation and assume homogeneous magnitude with component. We argue that this leads to a more flexible yet still interpretable parameterization.

One of the limitations of our work is that we assume spherical Gaussian in our spatial mixture models. For future work, we can use elliptical shape activations using general variance-covariance structures. Another limitation is the computational intensity. The present method took about 105 minutes on a MAC 2.7 GHz PowerPC G5. Although this is not excessive, with larger 3D images the compute time may be excessive. One avenue to explore would be to parallels the code. With the

proliferation of compute clusters, it would be a simple matter to schedule individuals to separate nodes. All population parameters would then be processed at each iteration after the scheduler returns results from the individuals. Another possible avenue would be a Variational Bayes approach (Attias, 2000), which would provide approximate posterior means and variances.

The flexibility of our model does make for some interpretive limitations. While most users of fMRI conceive of activation "loci" (x,y,z locations), in our model the population centers are random variables $\boldsymbol{\mu}$ whose distribution can't easily be summarized. We visualize this distribution with the Population Center Location posterior rate function image, inspecting for modes and assessing the spread about modes. At the subject level, the interpretation about individual activation centers is even more difficult. Both the "finite mixture model" and the "infinite mixture model" fit several individual component centers for one activation area. For example, in Figure 2.2 we plot the location of individual component centers at iteration 6000 for a representative sample of subjects (1, 3, 9 and 20) using the finite mixture Jump model. For subject 3, the model fits seven individual component centers for the right upper activation area. This improves performance in mixing and model fitting. However, the interpretation of the result as having as many as 11 individual activation centers may not be proper.

In future work we would like to address these shortcomings. In this Chapter we discuss possible solutions for the interpretation of individual activation centers in more details. When several activation centers are close together it is more reasonable to consider them as components of a single source of activation. The basic idea is to cluster these components centers and consider the cluster centers as the individual activation centers. This will add an additional level in the hierarchy of the Bayesian

model. We will focus on spatial cluster analysis based on Cox processes.

5.2 The Cox process

In this section we introduce the Cox process and discuss some of the Cox models that are related to our problem. The Cox process was studied by Cox (1955) under the name doubly stochastic Poisson process. For a point process X in $\mathbb{R}^d$ and a random measure M, X is said to be a Cox process driven by M if X given M follows a Poisson process with intensity measure M. In many applications M is specified by a nonnegative random intensity surface $\{Z(x) : x \in \mathbb{R}^d\}$ so that

$$M(A) = \int_A Z(x)dx, \text{ for a bounded region A} \in \mathbb{R}^d$$

Among the many approaches related to the Cox process, one of particular interest is the Neyman-Scott process (Neyman and Scott, 1958). In a Neyman-Scott process, the intensity surface is defined as a function of a set of cluster centers $C = \{c_1, c_2, ...\}$. That is, the clusters $X_1, X_2, ...$ are assumed to be independent Poisson processes with random intensity surface

$$Z(x) = \sum_i \alpha f(x - c_i)$$

where the cluster centers C follows a homogeneous Poisson process. Although the cluster centers C is modelled explicitly, in many applications C is not of interest and the estimation is based on the likelihood marginalized over C. The center location was first studied in Baddeley and van Lieshout (1993) and by Lawson (1993a). They used a modified Thomas function, which is given by

$$f(x \mid c) = \frac{\beta}{2\pi\sigma^2} e^{-(x-c)^2/2\sigma^2} \tag{5.1}$$

According to (4.1), the cluster components x follow a normal distribution with center c. There are several ways to modify the Neyman-Scott process. For example, we

can choose an arbitrary distribution for the cluster centers C. Sometimes cluster components may not have a center. For example a point represents noise or an outlier. This situation was considered in Dasgupta and Raftery (1998), van Lieshout (1995) and van Lieshout and Baddeley (1995). In their models a separate parameter ϵ was added to the intensity surface

$$Z(x) = \epsilon + \sum_i \alpha f(x - c_i)$$

and the noise or the outlier is modelled by a Poisson process of constant intensity ϵ. In the next section we will propose a Bayesian model based on these two extensions.

5.3 Methods

We start with the Bayesian model in Chapter II and add an additional level in the hierarchy by introducing a new parameter $\tilde{\eta} = \{\tilde{\eta}_{jh}\}$, $h = 1, \ldots, \tilde{c}_j$. In this section we will focus on the prior distribution of η_{jl}, $\tilde{\eta}_{jh}$ and $\tilde{c}_j$. To simplify the notation we drop the subject index j and work only with η_l, $\tilde{\eta}_h$, and $\tilde{c}$. We assume that given $\tilde{\eta}_h$, η_l follows a Poisson process with intensity surface

$$Z(\eta_l) = \epsilon + \sum_h \frac{\alpha_\eta}{2\pi\beta_\eta^2} e^{-(\eta_l - \tilde{\eta}_h)^2/2\beta_\eta^2}$$

This is the intensity surface Nyema-Scott processes with a modified Thomas function and extended to include a background intensity ϵ. We interpret the cluster center $\tilde{\eta}_h$ as the individual activation center and interpret $\tilde{c}$ as the number of individual activation centers. Write $M(A) = \int_A Z(x)dx$ where A is the brain area. It follows that the number of cluster components c follows a Poisson distribution with mean $M(A)$ and given c the location η_l are independently distributed with probability density $p(\eta_l) = Z(\eta_l)/M(A)$. We get the following priors:

$$p(\eta \mid \tilde{\eta}, \alpha_\eta, \beta_\eta, c, \tilde{c}) p(c \mid \tilde{\eta}, \tilde{c}) = \frac{e^{-M(A)} M(A)^c}{c!} \prod_{l=1}^{c} Z(\eta_l \mid \tilde{\eta})/M(A)$$

We further assume that the individual center location $\tilde{\eta}_h$ is from a normal mixture with each component associated with a particular population center μ_i.

$$f(\tilde{\eta}_h \mid \phi_i, \mu_i, \tau_i^2, c_p) = \sum_{i=1}^{c_p} \phi_i (2\pi\tau_i^2)^{-1} exp\ \{-\frac{(\tilde{\eta}_h - \mu_i)^T(\tilde{\eta}_h - \mu_i)}{2\tau_i^2}\}$$

This is the same prior used in chapter II for η.

With the new level, an additional reversible jump step for adding and deleting the individual activation center $\tilde{\eta}$ is required for posterior simulation. This will greatly increase the computational time. Future efforts can be made on looking for efficient ways of evaluating the integral in $M(A)$.

5.4 Conclusion

The main contributions of this work are as follows. A traditional voxel-based approach can only detect activated voxels for which the average signal is significant across subjects. However, subjects sometimes don't have, or only a small percentage of the total subjects have, overlapped activated voxels even though these activated voxels are close to each other and represent the same anatomical/functional area. Our models fit individual component centers for subject level activation areas and find population activation centers based on these individual component centers. We explicitly model the location of the population activation center with a separate parameter describing the variability of the individual component centers. Unlike the voxel-based approach our methods allow for variation in both signal intensity and activation location at the individual level. The location of the population activation centers are summarized using the posterior rate function. This provides better precision in the estimation of the population center location than the standard mass-univariate method.

APPENDICES

APPENDIX A

Details of the MCMC algorithm for the finite mixture model in Chapter II

In this Appendix we give the detailed MCMC simulation steps for the finite mixture model introduced in Chapter II. Current parameter values will be referred to by name and proposed parameters will be referred to by name superscripted with a prime ($'$). Each proposal is classified as a reversible jump proposal (RJ), a Metropolis-Hastings proposal (MH) or a Gibbs (Gibbs) proposal.

1. Birth of a population center: RJ.

   A population center birth is proposed with probability 0.5. Suppose there are c_p centers. Generate $\boldsymbol{\mu}'_{c_p+1}$, and $\tau^{2'}_{c_p+1}$ from their prior distributions:

   $$\boldsymbol{\mu}'_{c_p+1} \sim \text{uniform on} \cup_{j=1}^{J} A_j \quad \text{and} \quad \tau^{2'}_{c_p+1} \sim IG(3, \beta_\tau)$$

   Draw $\psi'_{c_p+1} \sim \text{Beta}(1, c_p)$ and re-scale the existing weights using $\psi'_i = \psi_i(1 - \psi'_{c_p+1})$ for $i = 1, \ldots, c_p$. Set $\boldsymbol{\mu}'_i = \boldsymbol{\mu}_i$ and $\tau^{2'}_i = \tau^2_i$ for $i = 1, \ldots, c_p$. Accept the new population center with probability

   $$\min \left\{ 1, \frac{5}{(c_p+1)^2} \prod_{j=1}^{J} \prod_{l=1}^{c_j} \frac{\sum_{i=1}^{c_p+1} \psi'_i \ \phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}'_i, \tau^{2'}_i \mathbf{I}_2)}{\sum_{i=1}^{c_p} \psi_i \ \phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \tau^2_i \mathbf{I}_2)} \right\}.$$

2. Death of a population center: RJ.

   The death of a population center is proposed with probability 0.5. Randomly

choose a center from the set {1,...,c_p}. Suppose it is j. Set $\psi_i' = \psi_i/(1-\psi_j)$ for all $i \in \{1,\ldots,c_p\} \setminus \{j\}$ and relabel the indices $1,\ldots,c_p-1$. Set $\boldsymbol{\mu}_i' = \boldsymbol{\mu}_i$ and $\tau_i^{2'} = \tau_i^2$ for $i = 1,\ldots,c_p-1$. Accept the death of this population center with probability

$$\min\left\{1, \frac{{c_p}^2}{5}\prod_{j=1}^{J}\prod_{l=1}^{c_j}\frac{\sum_{i=1}^{c_p-1}\psi_i'\phi_2(\boldsymbol{\eta}_{jl};\boldsymbol{\mu}_i',\tau_i^{2'}\mathbf{I}_2)}{\sum_{i=1}^{c_p}\psi_i\phi_2(\boldsymbol{\eta}_{jl};\boldsymbol{\mu}_i,\tau_i^2\mathbf{I}_2)}\right\}.$$

3. Birth of an individual component: RJ.

   For each subject j, a birth of a component is proposed with probability 0.5. Suppose there are c_j components for subject j. Proposed parameters are drawn from their prior distributions:

$$r_{j,\,c_j+1}^{2'} \sim IG(2\pi,\beta_r), \quad \sigma_{j,\,c_j+1}^{2'} \sim IG(3,\beta_\sigma)$$

$$Pr(z'_{j,\,c_j+1} = i) = \psi_i, \quad i = 1,\ldots,c_p$$

$$\left[\boldsymbol{\eta}'_{j,\,c_j+1} \mid z'_{j,\,c_j+1} = i\right] \sim N(\boldsymbol{\mu}_i,\tau_i^2), \quad \theta'_{j,\,c_j+1} \sim trunc_{(0,\infty)}N(\lambda_\theta,\sigma_\theta^2).$$

   Calculate p'_{jvl} according to equation 2.2 for all v, l and $j = 1,\ldots,c_j+1$. The acceptance probability is

$$\min\left\{1, 5(c_j+1)^{-2}(\text{likelihood ratio})\right\}.$$

   where the likelihood function is given in equation 2.1.

4. Death of an individual component: RJ.

   For each subject j, a death of a component is proposed with probability 0.5. Suppose there are c_j components for subject j. Randomly choose a component, j, from the set $\{1,...,c_j\}$. Calculate p'_{jvl} according to equation 2.2 for all v, l and $j = 1,\ldots,c_j-1$ and relabel the components $1,\ldots,c_j-1$. The acceptance probability is

$$\min\left\{1, {c_j}^2 5^{-1}(\text{likelihood ratio})\right\}.$$

where the likelihood function is given in equation 2.1.

5. Update $\boldsymbol{\mu}_i$ for all $i = 1, \ldots, c_p$: MH.

   Propose $\boldsymbol{\mu}_{\boldsymbol{i}}'$ from $N(\boldsymbol{\mu}_i, \sigma_\mu^2 \mathbf{I}_2)$. Set $\boldsymbol{\mu}_j' = \boldsymbol{\mu}_j$ for all $j \neq i$. Accept with probability

$$\min\left\{1, \frac{I(\boldsymbol{\mu}_i' \in \cup_{j=1}^J A_j)}{I(\boldsymbol{\mu}_i \in \cup_{j=1}^J A_j)} \prod_{j=1}^{J} \prod_{l=1}^{c_j} \frac{\sum_{i=1}^{c_p} \psi_i \phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i', \tau_i^2 \mathbf{I}_2)}{\sum_{i=1}^{c_p} \psi_i \phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \tau_i^2 \mathbf{I}_2)}\right\}.$$

6. Update $\boldsymbol{\eta}_{jl}$ for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$: MH.

   For subject j, component l, propose $\boldsymbol{\eta}_{jl}' \sim N(\boldsymbol{\eta}_{jl}, {\sigma_\eta}^2 \mathbf{I}_2)$. Calculate p_{jvl}' according to equation 2.2. Set $\boldsymbol{\eta}_{kl}' = \boldsymbol{\eta}_{kl}$ for all $k \neq j$ and all $l = 1, \ldots, c_j$. Accept with probability

$$\min\left\{1, (\text{likelihood ratio}) \times \frac{\sum_{i=1}^{c_p} \psi_i \phi_2(\boldsymbol{\eta}_{jl}'; \boldsymbol{\mu}_i, \tau_i^2 \mathbf{I}_2)}{\sum_{i=1}^{c_p} \psi_i \phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \tau_i^2 \mathbf{I}_2)}\right\}.$$

   where the likelihood function is given in equation 2.1.

7. Update τ_i^2 for all $i = 1, \ldots, c_p$: Gibbs.

   Let D_i be the set of individual components with $z_{jl} = i$ and let N_{D_i} be the number of individual components in D_i.

   Draw $\tau_i^2 \sim IG(N_{D_i} + 3, 0.5 \sum_{jl \in D_i} (\boldsymbol{\eta}_{jl} - \boldsymbol{\mu}_i)^T (\boldsymbol{\eta}_{jl} - \boldsymbol{\mu}_i) + \beta_\tau)$.

8. Update r_{jl}^2 for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$: MH.

   For subject j, component l, propose $\log r_{jl}^{2\prime} \sim N(\log r_{jl}^2, \sigma_r^2)$. Calculate p_{jvl}' according to equation 2.2 for all v. Accept with probability

$$\min\left\{1, (r_{jl}^2 / r_{jl}^{2\prime})^{2\pi} \exp[\beta_r (r_{jl}^{-2} - r_{jl}^{-2\prime})] (\text{likelihood ratio})\right\}.$$

   where the likelihood function is given in equation 2.1.

9. Update $\psi_1, \ldots, \psi_{c_p}$: Gibbs.

   Draw $(\psi_1, \psi_2, \ldots, \psi_{c_p})$ from $D(1 + N_{D_1}, ..., 1 + N_{D_{c_p}})$.

10. Update z_{jl} for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$: Gibbs.

    Draw z_{jl} according to $\Pr(z_{jl} = i \mid .) \propto \psi_i \phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \tau_i^2 \mathbf{I}_2)$.

11. Update ω_{jv} for all $j = 1, \ldots, J$ and all pixels v: Gibbs.

    Draw ω_{jv} from $\Pr(\omega_{jv} = l \mid .) \propto p_{jvl} \phi(y_{jv}; \theta_{jl}, \sigma_{jl}^2)$.

12. Update σ_{jl}^2 for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$: Gibbs.

    Let G_{jl} denote the set of pixels with $\omega_{jv} = l$ for subject j. Let $N_{G_{jl}}$ be the number of voxels in G_{jl}. Draw $\sigma_{jl}^2 \sim IG(0.5 N_{G_{jl}} + 3, 0.5 \sum_{v \in G_{jl}} (y_{jv} - \theta_{jl})^2 + \beta_\sigma)$.

13. Update σ_0^2: Gibbs.

    Let G_0 denote the set of pixels with $\omega_{jv} = 0$ for subject j.

    Draw $\sigma_0^2 \sim IG\left(0.5 \sum_{j=1}^J N_{G_{j0}} + 10^{-3}, 0.5 \sum_{j=1}^J \sum_{v \in G_{j0}} (y_{jv} - \theta_0)^2 + 10^{-3}\right)$.

14. Update θ_{jl} for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$: Gibbs.

    Draw $\theta_{jl} \sim N(mv, v)$, where

    $$m = \sigma_\theta^{-2} \lambda_\theta + \sigma_{jl}^{-2} \sum_{v \in G_{jl}} y_{jv} \quad \text{and} \quad v = \left(\sigma_{jl}^{-2} N_{G_{jl}} + \sigma_\theta^{-2}\right)^{-1}.$$

    Accept if $\theta'_{jl} > 0$.

15. Update θ_0: Gibbs.

    Draw $\theta_0 \sim N(mv, v)$, where

    $$m = \sigma_0^{-2} \sum_{j=1}^J \sum_{v \in G_{j0}} y_{jv} \quad \text{and} \quad v = \left(\sigma_0^{-2} \sum_{j=1}^J N_{G_{j0}} + 1\right)^{-1}.$$

16. Update β_σ: Gibbs.

    Draw $\beta_\sigma \sim Ga(3 \sum_{j=1}^J c_j + 0.01, \sum_{j=1}^J \sum_{l=1}^{c_j} \sigma_{jl}^{-2} + 0.01)$. where $Ga(\kappa, \lambda) = \frac{\lambda^\kappa}{\Gamma(\kappa)} x^{\kappa-1} e^{-\lambda x}$ is the gamma distribution

17. Update β_τ: Gibbs.

    Draw $\beta_\tau \sim Ga(3c_p + 0.01, \sum_{i=1}^{c_p} \tau_i^{-2} + 0.01)$.

18. Update β_r: Gibbs.

    Draw $\beta_r \sim Ga(2\pi \sum_{j=1}^{J} c_j + 0.01, \sum_{j=1}^{J} \sum_{l=1}^{c_j} r_{jl}^{-2} + 0.01)$.

19. Update λ_θ: Gibbs.

    Draw $\lambda_\theta \sim N(mv, v)$, where

$$m = \sigma_\theta^{-2} \sum_{j=1}^{J} \sum_{l=1}^{c_j} \theta_{jl} + (35)1e^{-8} \quad \text{and} \quad v = (\sigma_\theta^{-2} \sum_{j=1}^{J} c_j + 1e^{-8})^{-1}.$$

20. Update σ_θ^2: Gibbs.

    Draw $\sigma_\theta^2 \sim IG(0.5 \sum_{j=1}^{J} c_j + 0.01, 0.5 \sum_{j=1}^{J} \sum_{l=1}^{c_j} (\theta_{jl} - \lambda_\theta)^2 + 0.01)$.

APPENDIX B

Details of the MCMC algorithm for the infinite mixture model in Chapter III

In this Appendix we give the detailed MCMC simulation steps for the infinite mixture model introduced in Chapter III. Current parameter values will be referred to by name and proposed parameters will be referred to by name superscripted with a prime ($'$).

1. Birth of an individual component:

   For each subject j, a birth of a component is proposed with probability 0.5. Suppose there are c_p population centers and there are c_j components for subject j. Let D_i be the set of individual components with $z_{jl} = i$ and let N_{D_i} be the number of individual components in D_i. Draw $z'_{j,\,c_j+1}$ according to the following probabilities:

$$P(z'_{j,\,c_j+1} = i \mid z_{j1}, ..., z_{j,c_j}) = \begin{cases} N_{D_i}/(\alpha_0 + \sum_{j=1}^{J} c_j) & \text{for } 1 \le i \le c_p, \\ \alpha_0/(\alpha_0 + \sum_{j=1}^{J} c_j) & \text{for } i = c_p + 1. \end{cases}$$

   If $z'_{j,\,c_j+1} = c_p + 1$, draw $(\boldsymbol{\mu}_{c_p+1}, \boldsymbol{\Sigma}_{c_p+1})$ from G_0. Propose other parameters from their prior distributions:

$$r^{2'}_{j,\,c_j+1} \sim IG(2\pi, \beta_r), \quad \sigma^{2'}_{j,\,c_j+1} \sim IG(3, \beta_\sigma)$$

$$\left[\boldsymbol{\eta}'_{j,\,c_j+1} \mid z_{j,\,c_j+1} = i\right] \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i), \quad \theta'_{j,\,c_j+1} \sim trunc_{(0,\infty)} N(\lambda_\theta, \sigma_\theta^2).$$

Calculate p'_{jvl} according to equation 2.2 for all v, l and $j = 1, \ldots, c_j + 1$. The acceptance probability is

$$\min\left\{1, 5(c_j+1)^{-2}(\text{likelihood ratio})\right\}.$$

where the likelihood function is given in equation 2.1. If this birth is accepted when $z'_{j,\,c_j+1} = c_p + 1$, the new population center with parameter $(\boldsymbol{\mu}_{c_p+1}, \boldsymbol{\Sigma}_{c_p+1})$ should be inserted and the number of population centers increases from c_p to $c_p + 1$.

2. Death of an individual component:

   For each subject j, a death of a component is proposed with probability 0.5. Suppose there are c_p population centers and there are c_j components for subject j. Randomly choose a component, j, from the set $\{1, ..., c_j\}$. Calculate p'_{jvl} according to equation 2.2 for all v, l and $j = 1, \ldots, c_j - 1$ and relabel the components $1, \ldots, c_j - 1$. The acceptance probability is

$$\min\left\{1, {c_j}^2 5^{-1}(\text{likelihood ratio})\right\}.$$

   where the likelihood function is given in equation 2.1. If this proposal is accepted when $N_{D_{jl}} = 1$, the empty population center with parameter $(\boldsymbol{\mu}_{z_{jl}}, \boldsymbol{\Sigma}_{z_{jl}})$ should be removed and the number of population centers reduces to $c_p - 1$.

3. Update z_{jl} for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$:

   Let D_i be the set of individual components with $z_{jl} = i$ and let N_{D_i} be the number of individual components in D_i. If $N_{D_{z_{jl}}} > 1$, draw $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ from G_0 for which $c_p < i \leq c_p + m$, where m is the number of auxiliary components. Draw

a new z'_{jl} using the following probabilities:

$$P(z'_{jl} = i \mid z_{-jl}, \boldsymbol{\eta}_{jl}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \begin{cases} N_{D_i}\phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) & \text{for } 1 \leq i \leq c_p, \\ \alpha_0/M\phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) & \text{for } c_p < i \leq c_p + m. \end{cases}$$

If $z'_{jl} = i$ for $c_p < i \leq c_p + m$, add a new population center $(\boldsymbol{\mu}_{c_p+1}, \boldsymbol{\Sigma}_{c_p+1})$= $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$. If $N_{D_{z_{jl}}} = 1$, draw $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$ from G_0 for which $c_p < i \leq c_p + m$. Draw a new z'_{jl} using the following probabilities:

$$P(z'_{jl} = i \mid z_{-jl}, \boldsymbol{\eta}_{jl}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \propto \begin{cases} N_{D_i}\phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) & \text{for } 1 \leq i \leq c_p \text{ and } i \neq z_{jl}, \\ \alpha_0/M\phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i) & \text{for } i = z_{jl} \text{ or } c_p < i \leq c_p + m. \end{cases}$$

If $z'_{jl} \neq z_{jl}$ remove population center z_{jl} and the number of population center reduces to $c_p - 1$. If $z'_{jl} = i$ for $c_p < i \leq c_p + m$, add a new population center with parameters $(\boldsymbol{\mu}_{c_p}, \boldsymbol{\Sigma}_{c_p})$= $(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)$.

4. Update $\boldsymbol{\mu}_i$ for all $i = 1, \ldots, c_p$:

   Let D_i be the set of individual components with $z_{jl} = i$. Propose $\boldsymbol{\mu}'_i$ from $N(\boldsymbol{\mu}_i, \sigma^2_\mu \mathbf{I}_2)$. Set $\boldsymbol{\mu}'_j = \boldsymbol{\mu}_j$ for all $j \neq i$. Accept with probability

   $$\min\left\{1, \frac{I(\boldsymbol{\mu}'_i \in \cup_{j=1}^J A_j)}{I(\boldsymbol{\mu}_i \in \cup_{j=1}^J A_j)} \prod_{jl \in D_i} \frac{\phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}'_i, \boldsymbol{\Sigma}_i)}{\phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i)}\right\}.$$

5. Update $\boldsymbol{\Sigma}_i$ for all $i = 1, \ldots, c_p$:

   Let D_i be the set of individual components with $z_{jl} = i$ and let N_{D_i} be the number of individual components in D_i.

   Draw $\boldsymbol{\Sigma}_i \sim IW_{N_{D_i}+\nu}\left(\left(\boldsymbol{S} + \sum_{jl \in D_i} (\boldsymbol{\eta}_{jl} - \boldsymbol{\mu}_i)(\boldsymbol{\eta}_{jl} - \boldsymbol{\mu}_i)^T\right)^{-1}\right)$

6. Update $\boldsymbol{\eta}_{jl}$ for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$:

   For subject j, component l, propose $\boldsymbol{\eta}'_{jl} \sim N(\boldsymbol{\eta}_{jl}, {\sigma_\eta}^2 \mathbf{I}_2)$. Calculate p'_{jvl}. Set

$\boldsymbol{\eta}'_{kl} = \boldsymbol{\eta}_{kl}$ for all $k \neq j$ and all $l = 1, \ldots, c_j$. Accept with probability

$$\min\left\{1, (\text{likelihood ratio}) \times \frac{\phi_2(\boldsymbol{\eta}'_{jl}; \boldsymbol{\mu}_{z_{jl}}, \boldsymbol{\Sigma}_{z_{jl}})}{\phi_2(\boldsymbol{\eta}_{jl}; \boldsymbol{\mu}_{z_{jl}}, \boldsymbol{\Sigma}_{z_{jl}})}\right\}.$$

7. Update r^2_{jl} for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$:

   For subject j, component l, propose $\log r^{2\prime}_{jl} \sim N(\log r^2_{jl}, \sigma^2_r)$. Calculate p'_{jvl} for all v. Accept with probability

$$\min\left\{1, (r^2_{jl}/r^{2'}_{jl})^{2\pi} \exp[\beta_r(r^{-2}_{jl} - r^{-2'}_{jl})](\text{likelihood ratio})\right\}.$$

8. Update ω_{jv} for all $j = 1, \ldots, J$ and all pixels v:

   Draw ω_{jv} from $\Pr(\omega_{jv} = l \mid .) \propto p_{jvl}\phi(y_{jv}; \theta_{jl}, \sigma^2_{jl})$.

9. Update σ^2_{jl} for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$:

   Let B_{jl} denote the set of pixels with $\omega_{jv} = l$ for subject j. Let $N_{B_{jl}}$ be the number of voxels in B_{jl}. Draw $\sigma^2_{jl} \sim IG(0.5N_{B_{jl}} + 3, 0.5\sum_{v\in B_{jl}}(y_{jv} - \theta_{jl})^2 + \beta_\sigma)$.

10. Update σ^2_0 Let B_0 denote the set of pixels with $\omega_{jv} = 0$ and let N_{B_0} be the number of voxels in B_0.

    Draw $\sigma^2_0 \sim IG\left(0.5\sum_{j=1}^J N_{B_0} + 10^{-3}, 0.5\sum_{j=1}^J \sum_{v\in B_0}(y_{jv} - \theta_0)^2 + 10^{-3}\right)$.

11. Update θ_{jl} for all $j = 1, \ldots, J$ and $l = 1, \ldots, c_j$:

    Draw $\theta_{jl} \sim N(mv, v)$, where $m = \sigma_\theta^{-2}\lambda_\theta + \sigma_{jl}^{-2}\sum_{v\in G_{jl}} y_{jv}$

    and $v = \left(\sigma_{jl}^{-2}N_{G_{jl}} + \sigma_\theta^{-2}\right)^{-1}$. Accept if $\theta'_{jl} > 0$.

12. Update θ_0. Draw $\theta_0 \sim N(mv, v)$, where

$$m = \sigma_0^{-2}\sum_{j=1}^J \sum_{v\in B_0} y_{jv} \quad \text{and} \quad v = \left(\sigma_0^{-2}\sum_{j=1}^J N_{B_0} + 1\right)^{-1}.$$

13. Update β_σ. Draw $\beta_\sigma \sim Ga(3\sum_{j=1}^J c_j + 0.01, \sum_{j=1}^J \sum_{l=1}^{c_j} \sigma_{jl}^{-2} + 0.01)$., where $Ga(\kappa, \lambda) = \frac{\lambda^\kappa}{\Gamma(\kappa)}x^{\kappa-1}e^{-\lambda x}$ is the gamma distribution.

14. Update $\boldsymbol{S}$. Draw $\boldsymbol{S} \sim W_{\nu c_p + f}\left(\left(\boldsymbol{T}^{-1} + \sum_{i=1}^{c_p} \boldsymbol{\Sigma}_i^{-1}\right)^{-1}\right)$.

15. Update β_r. Draw $\beta_r \sim Ga(2\pi \sum_{j=1}^{J} c_j + 0.01, \sum_{j=1}^{J} \sum_{l=1}^{c_j} r_{jl}^{-2} + 0.01)$.

16. Update λ_θ. Draw $\lambda_\theta \sim N(mv, v)$, where

$$m = \sigma_\theta^{-2} \sum_{j=1}^{J} \sum_{l=1}^{c_j} \theta_{jl} + (35)1e^{-8} \quad \text{and} \quad v = (\sigma_\theta^{-2} \sum_{j=1}^{J} c_j + 1e^{-8})^{-1}.$$

17. Update σ_θ^2. Draw $\sigma_\theta^2 \sim IG(0.5 \sum_{j=1}^{J} c_j + 0.01, 0.5 \sum_{j=1}^{J} \sum_{l=1}^{c_j} (\theta_{jl} - \lambda_\theta)^2 + 0.01)$.

18. Update α_0.

    This update is based on the method proposed by Escobar and West 1995[10].

    Draw $\alpha_0 \sim \begin{cases} Ga(c_p + 1, 1 - ln(\beta_0)) & \text{with Prob. } \propto c_p, \\ Ga(c_p, 1 - ln(\beta_0)) & \text{with Prob. } \propto (1 - ln(\beta_0)) \sum_{j=1}^{J} c_j + c_p. \end{cases}$

    where β_0 is from Beta$(\alpha_0 + 1, \sum_{j=1}^{J} c_j)$

APPENDIX C

Real data results of all 21 subjects for the finite mixture model in Chapter II

In this Appendix we show the real data results for the finite mixture model introduced in Chapter II. Figures C.1 to C.7 show the source data, the marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$ and the center location of individual components at iteration 6000. Figure C.8 shows the trace plots of the number of population centers and the center of individual components. Figure C.9 shows the posterior distributions of the number of population centers and the center of individual components.

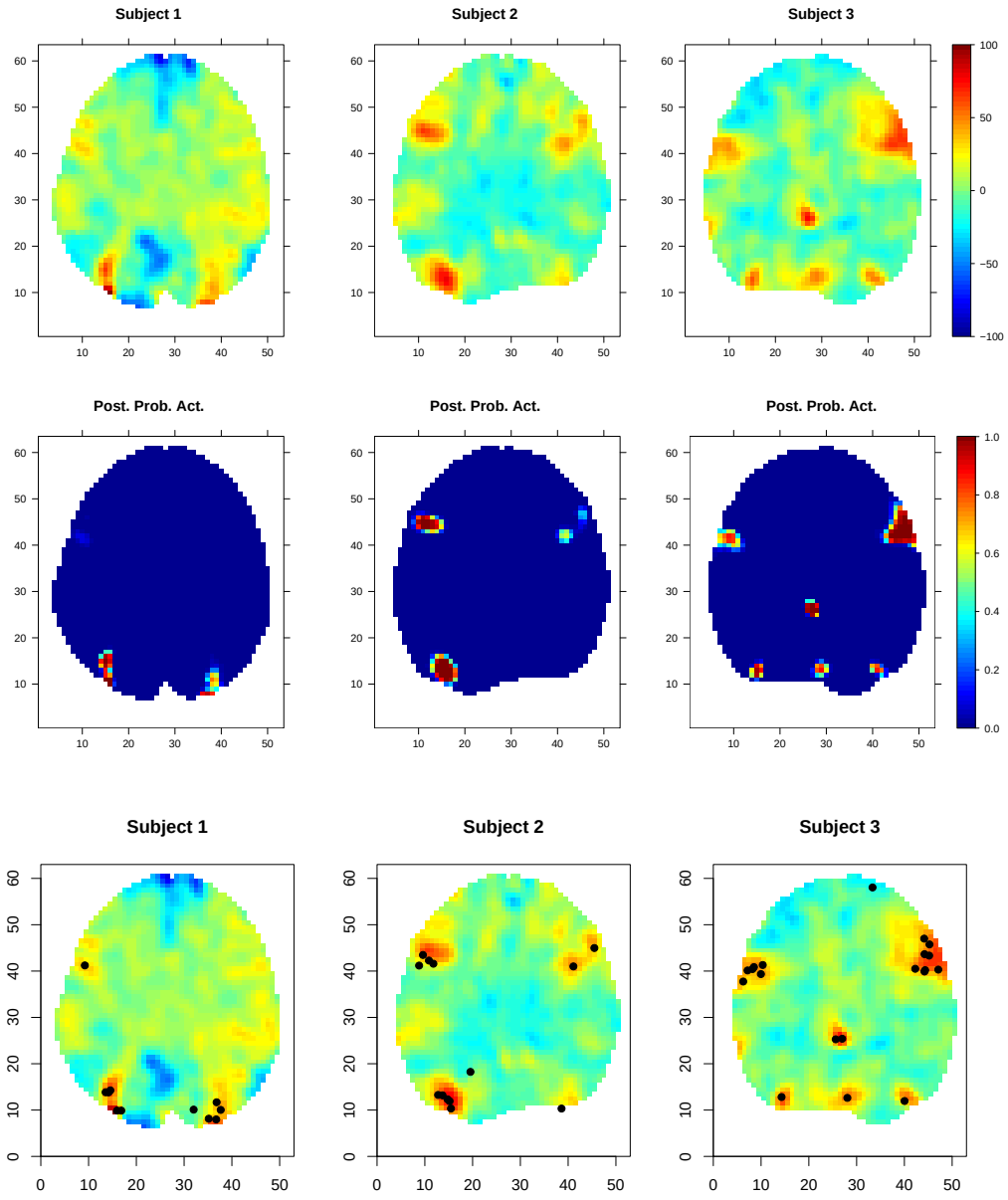


Figure C.1: Top row: The intensity data for subject 1-3. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

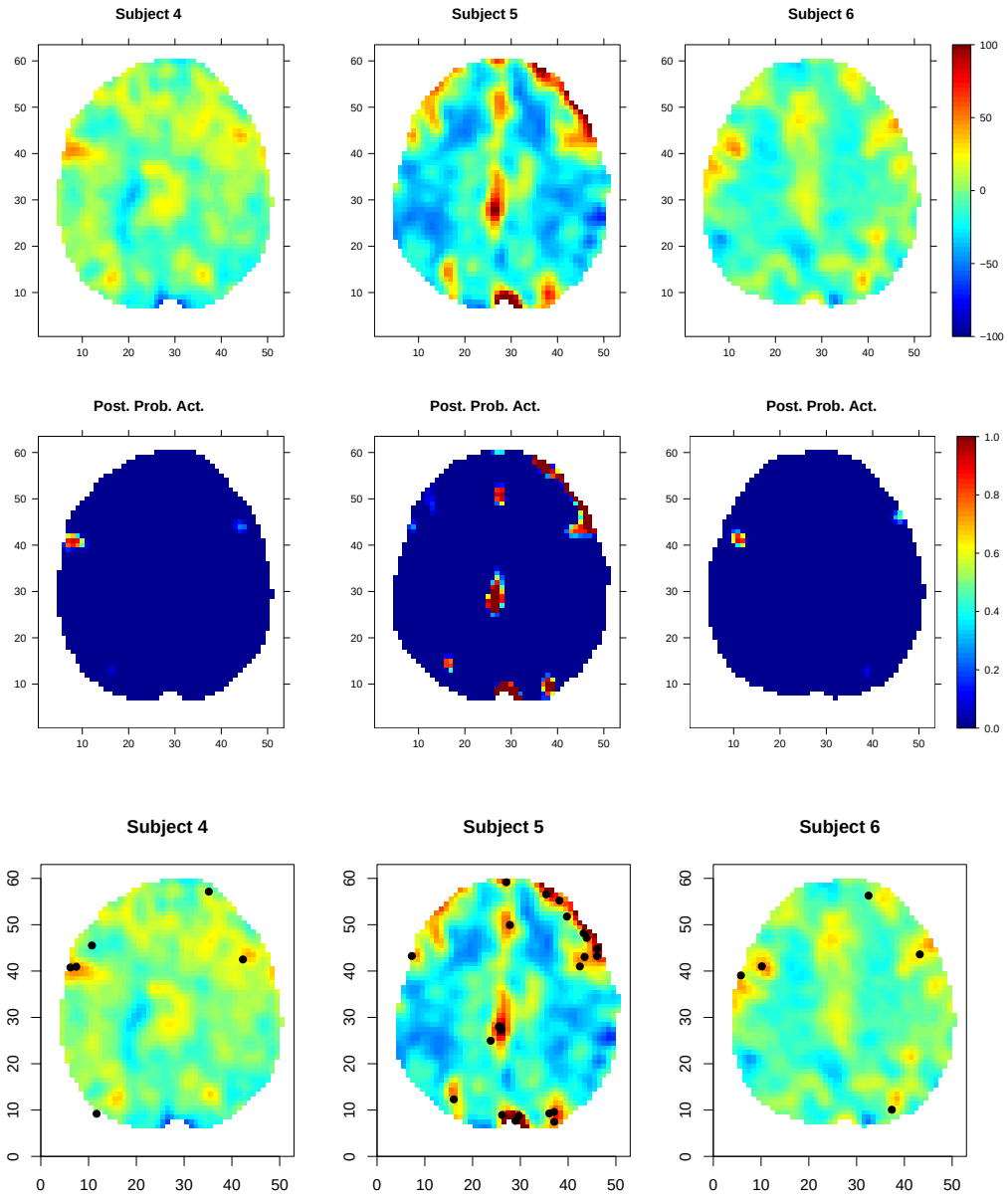


Figure C.2: Top row: The intensity data for subject 4-6. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

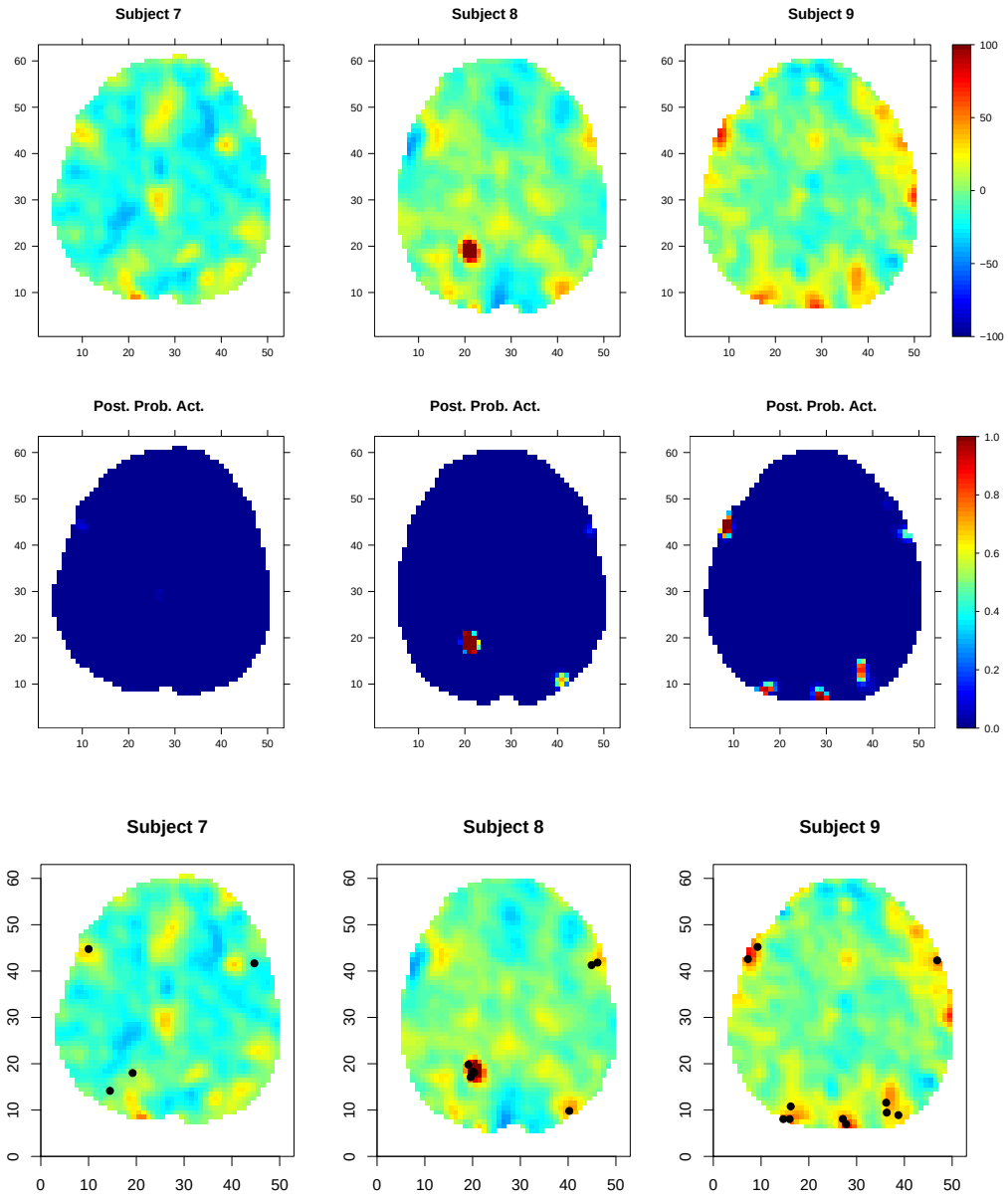


Figure C.3: Top row: The intensity data for subject 7-9. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

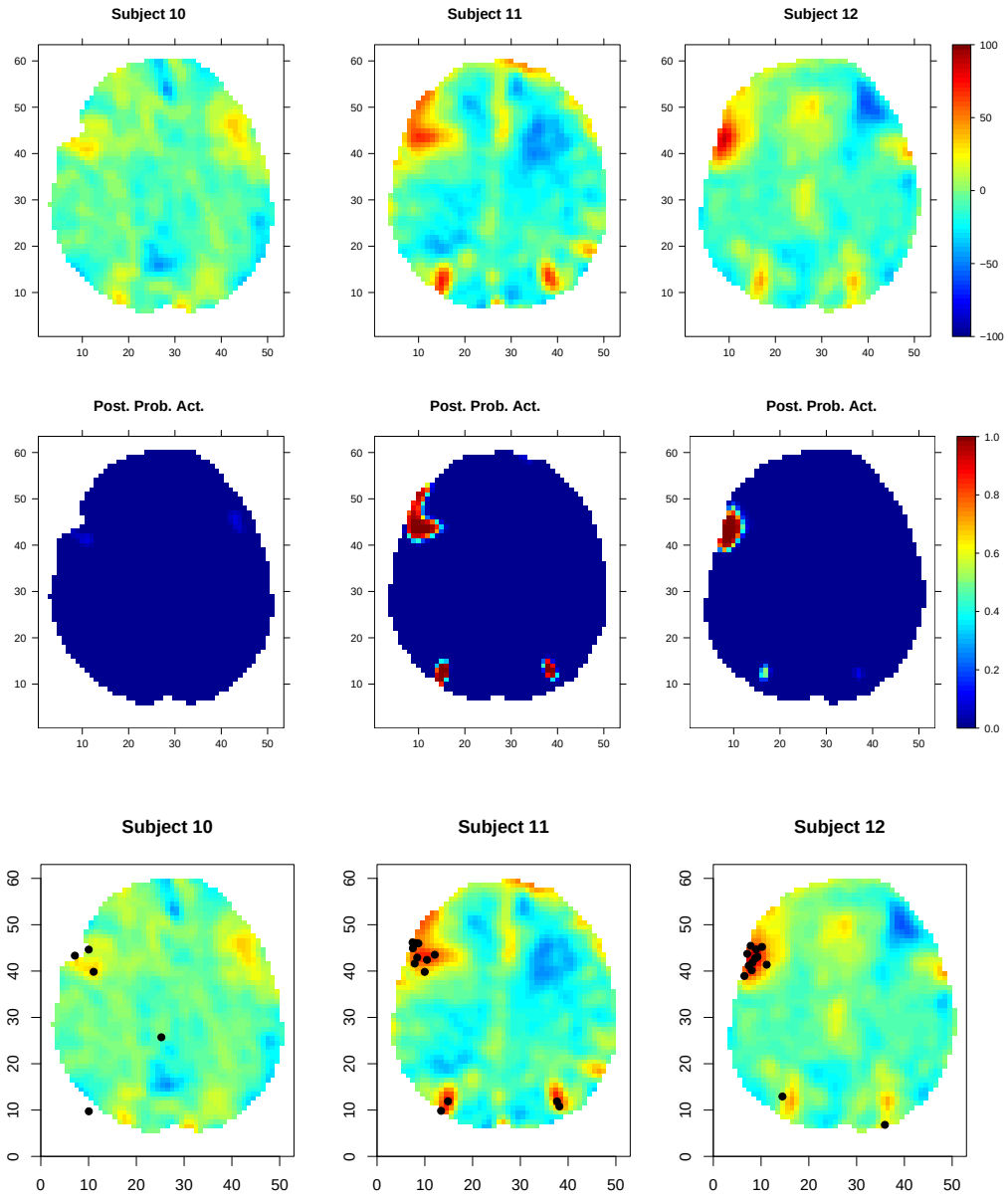


Figure C.4: Top row: The intensity data for subject 10-12. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

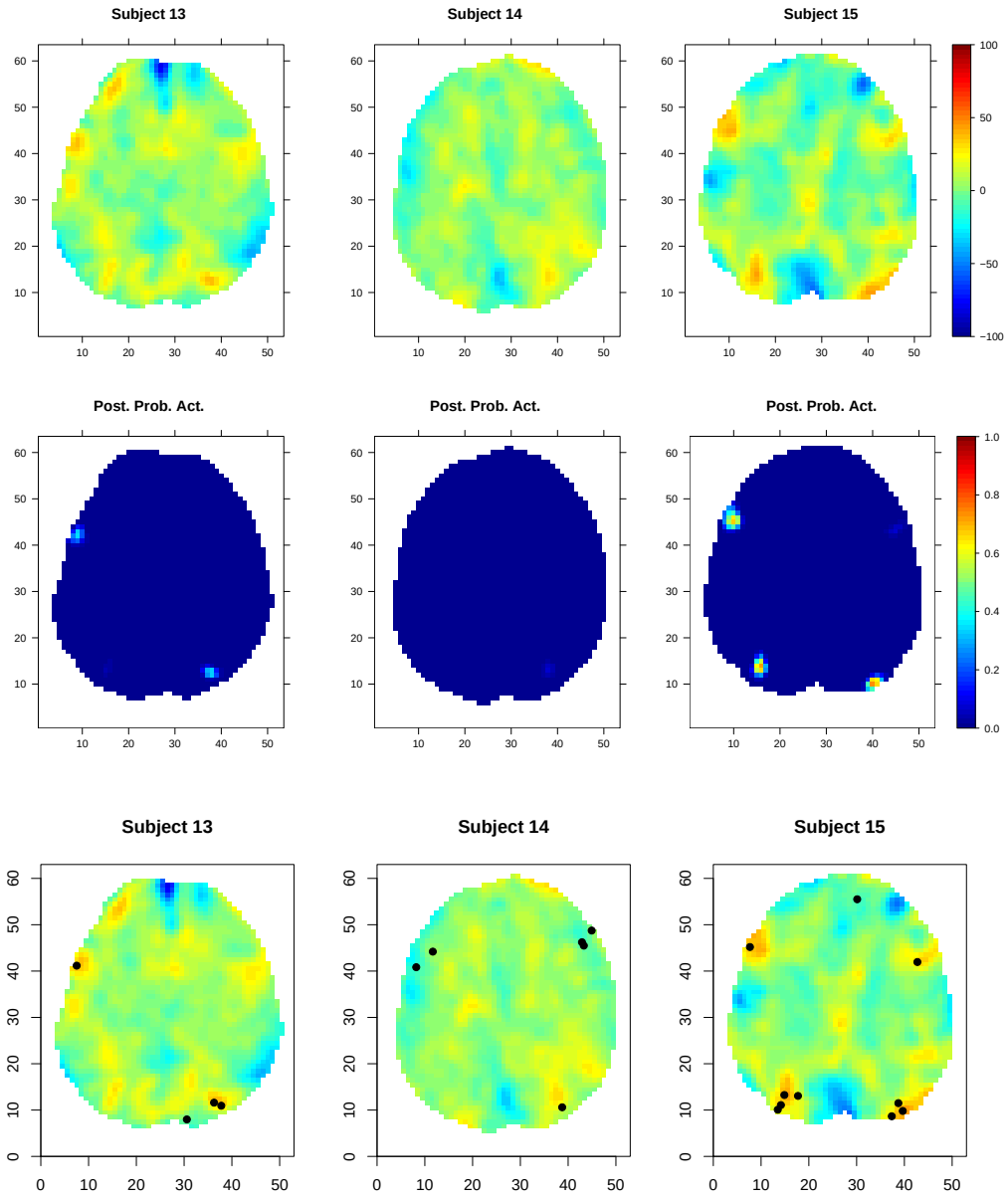


Figure C.5: Top row: The intensity data for subject 13-15. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

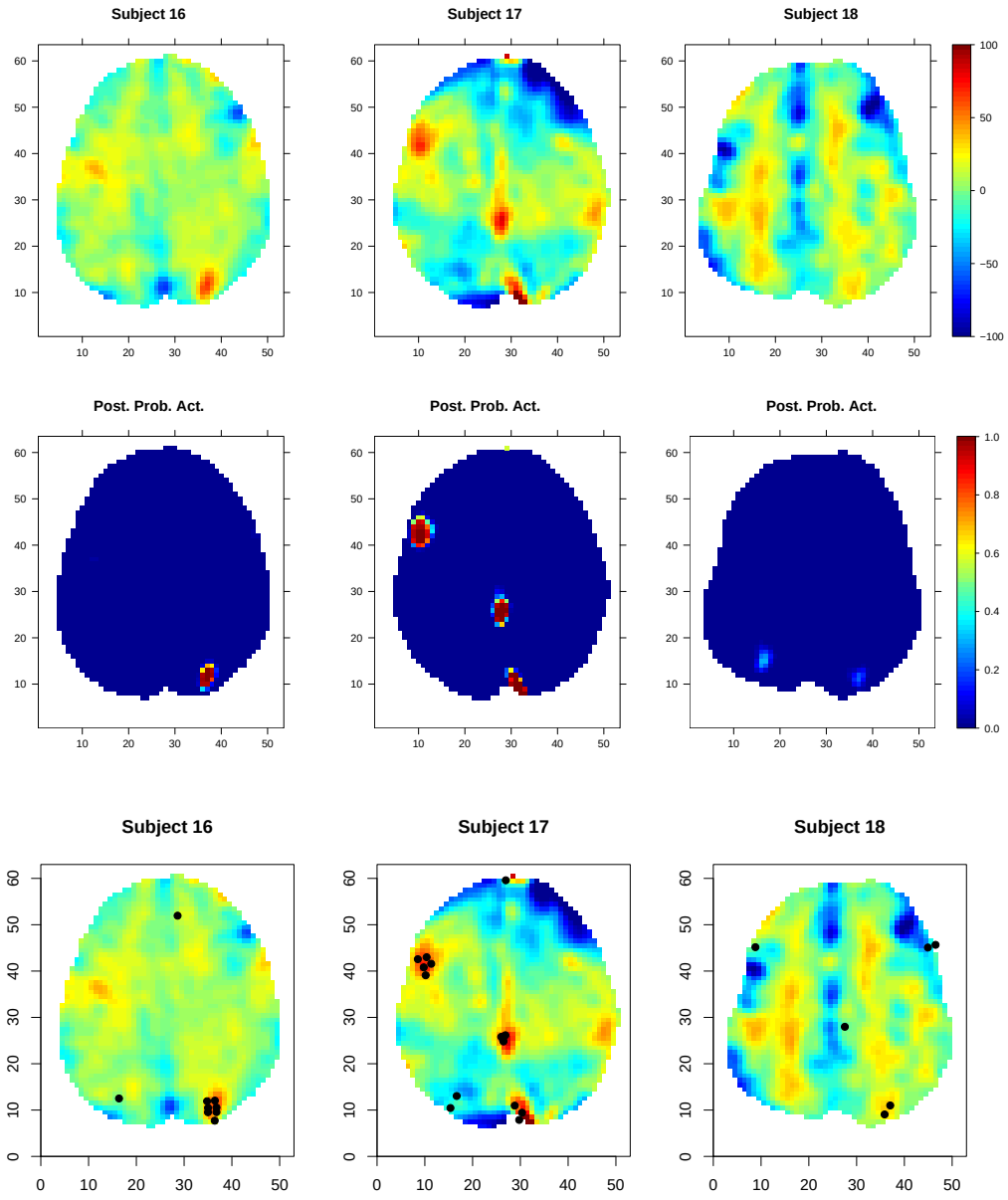


Figure C.6: Top row: The intensity data for subject 16-18. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

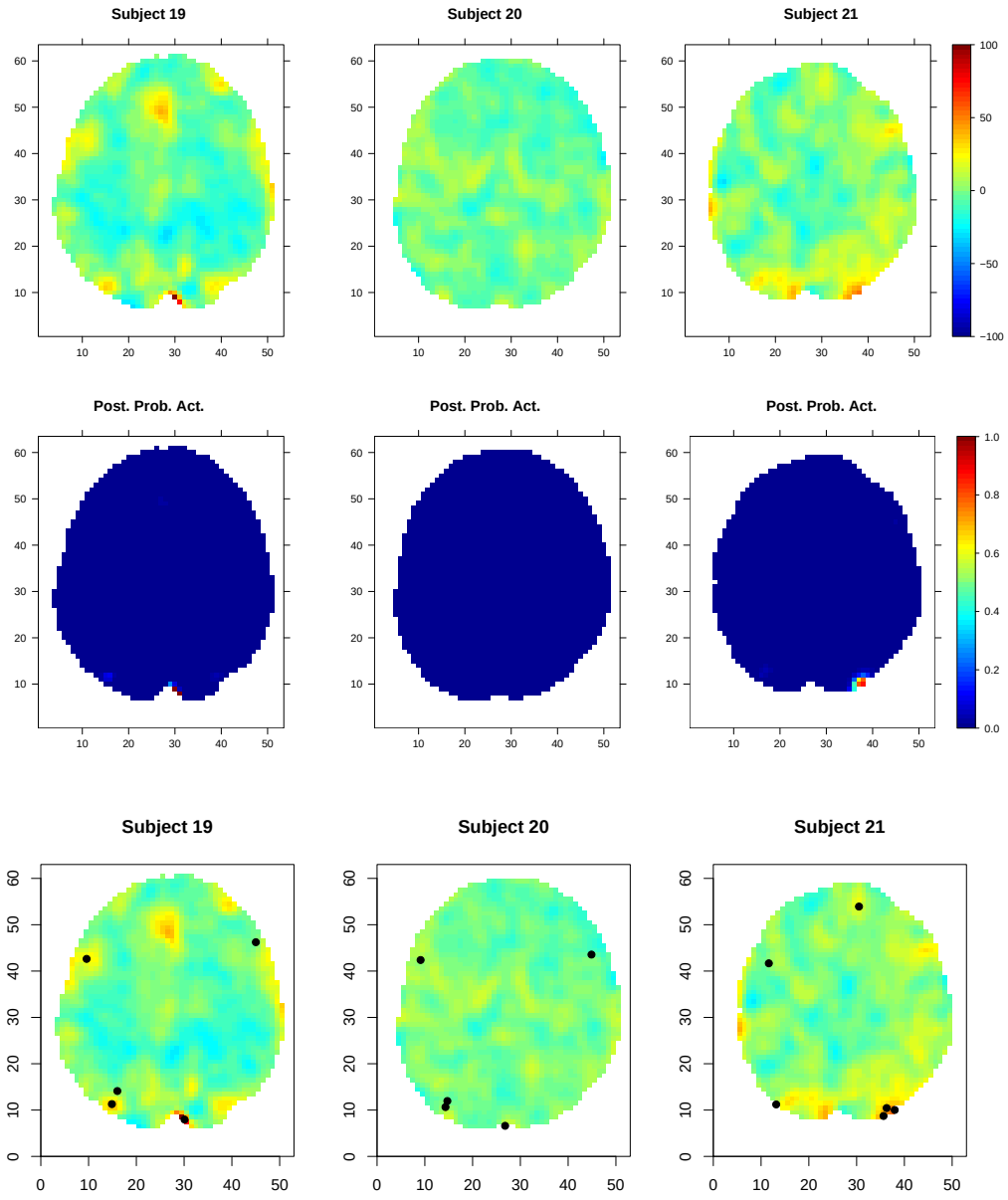


Figure C.7: Top row: The intensity data for subject 19-21. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

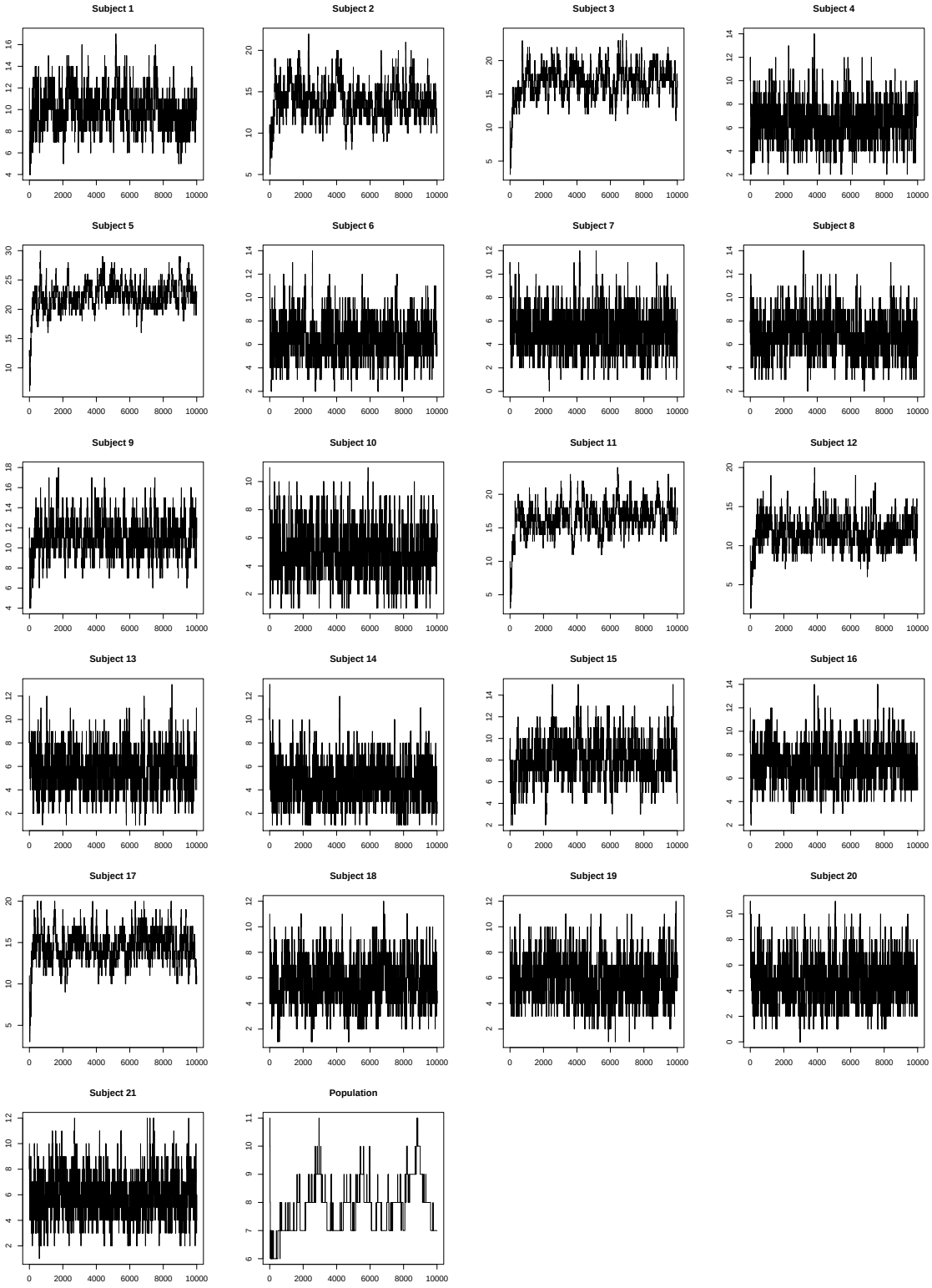


Figure C.8: Trace plot of the number of population centers and the center of individual components including the 5000 burn-in iterations

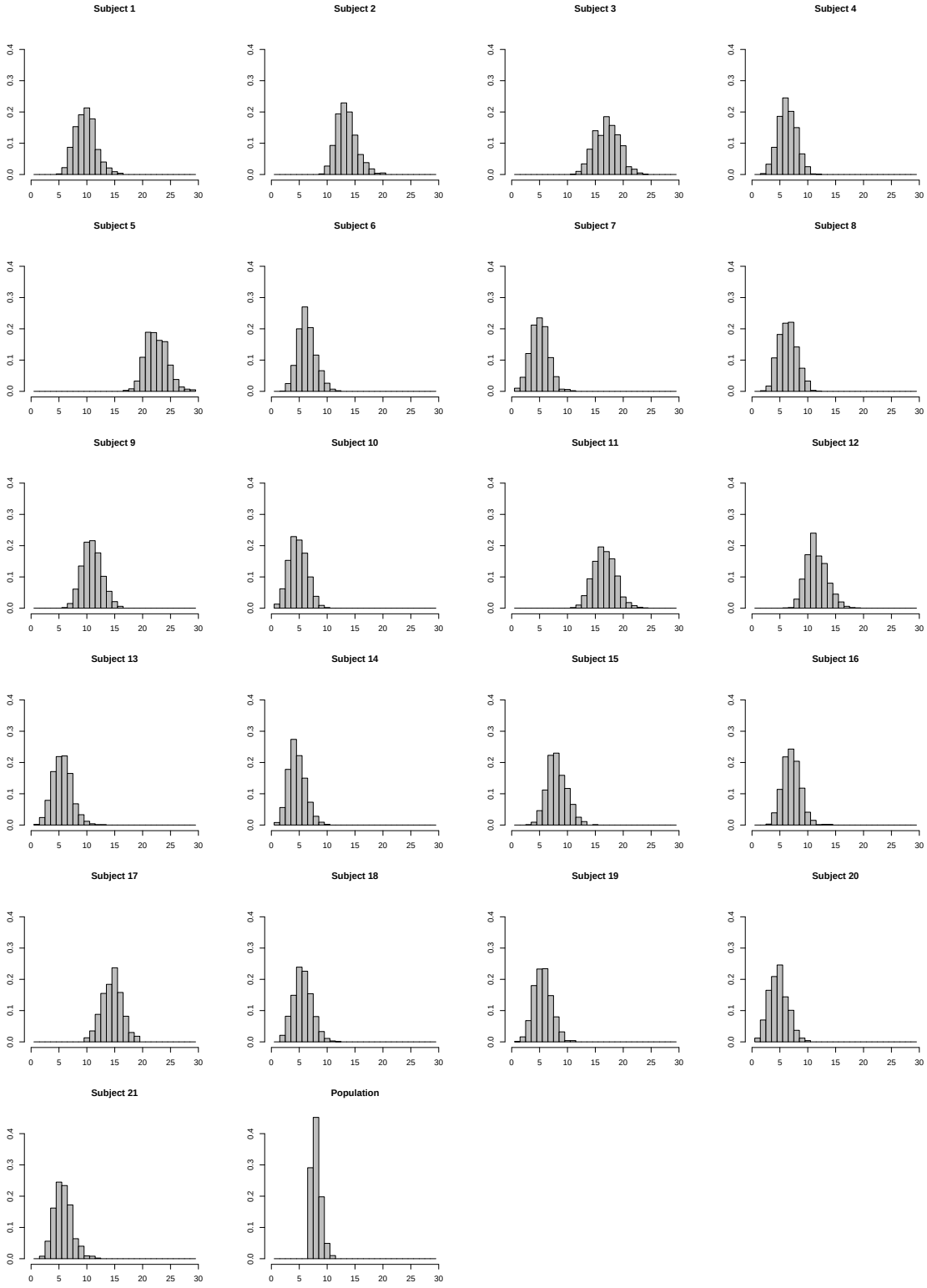


Figure C.9: Posterior distributions of the number of population centers and the center of individual components

APPENDIX D

Real data results of all 21 subjects for the infinite mixture model in Chapter III

In this Appendix we show the real data results for the infinite mixture model introduced in Chapter III. Figures D.1 to D.7 show the source data, the marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$ and the center location of individual components at iteration 6000. Figure D.8 shows the trace plots of the number of population centers and the center of individual components. Figure D.9 shows the posterior distributions of the number of population centers and the center of individual components.

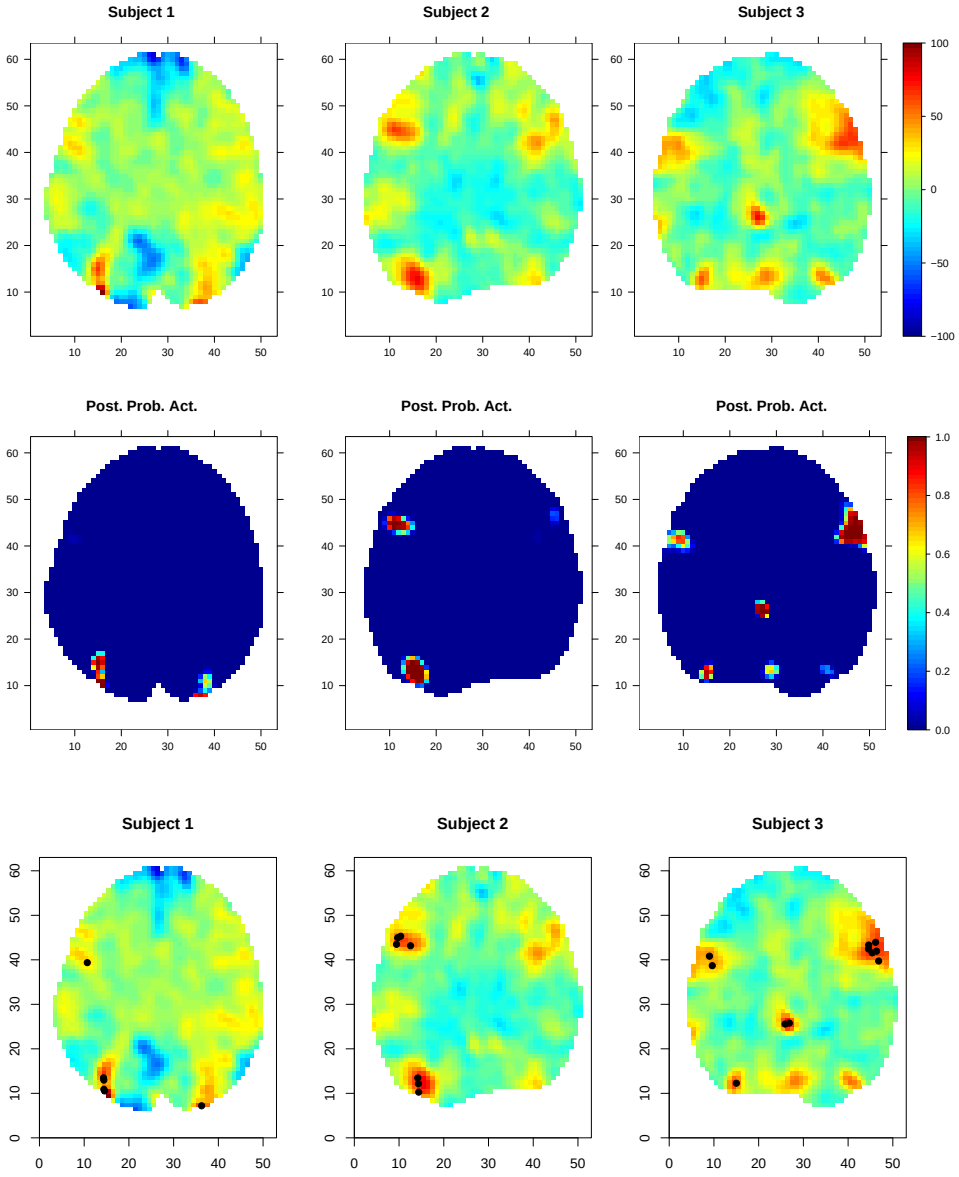


Figure D.1: Top row: The intensity data for subject 1-3. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

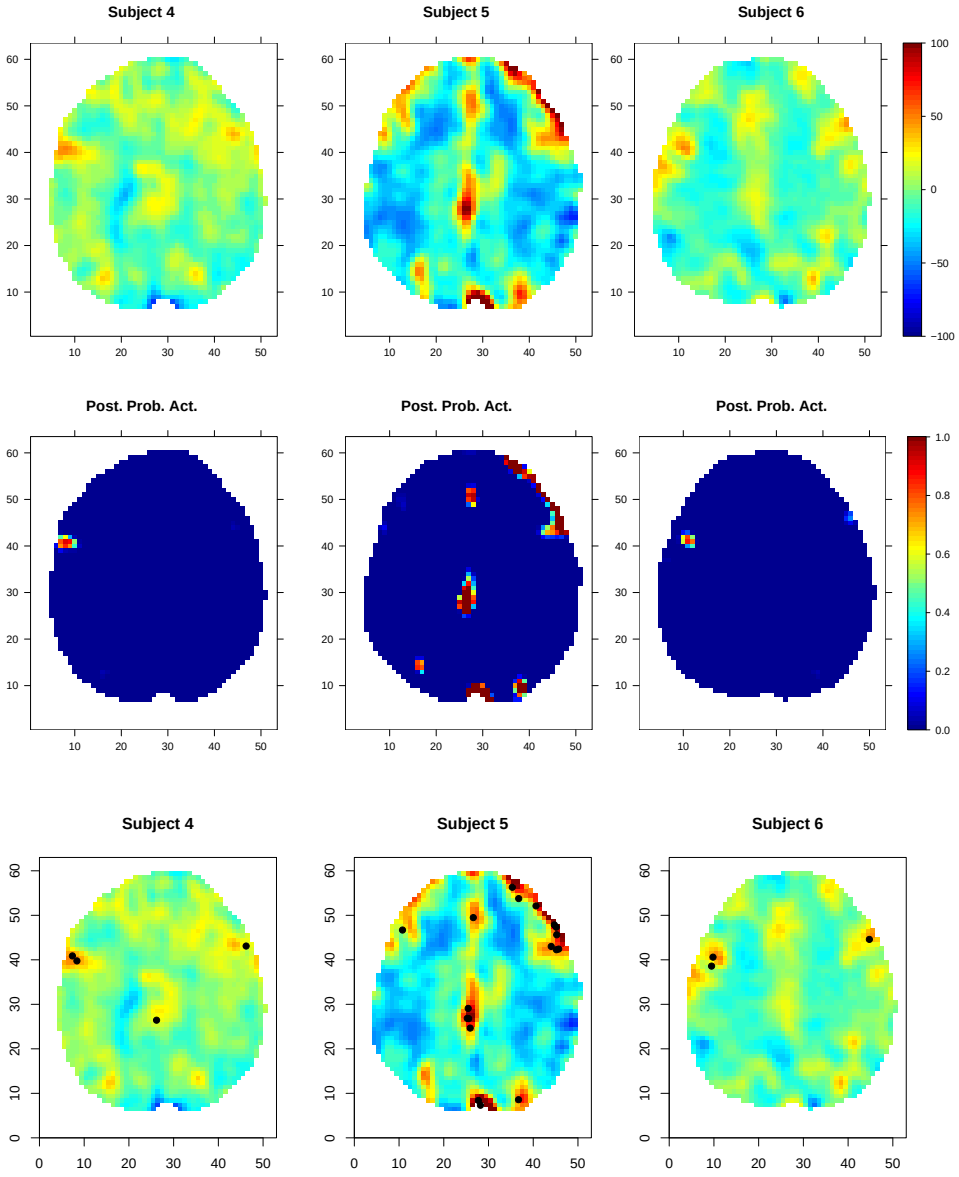


Figure D.2: Top row: The intensity data for subject 4-6. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

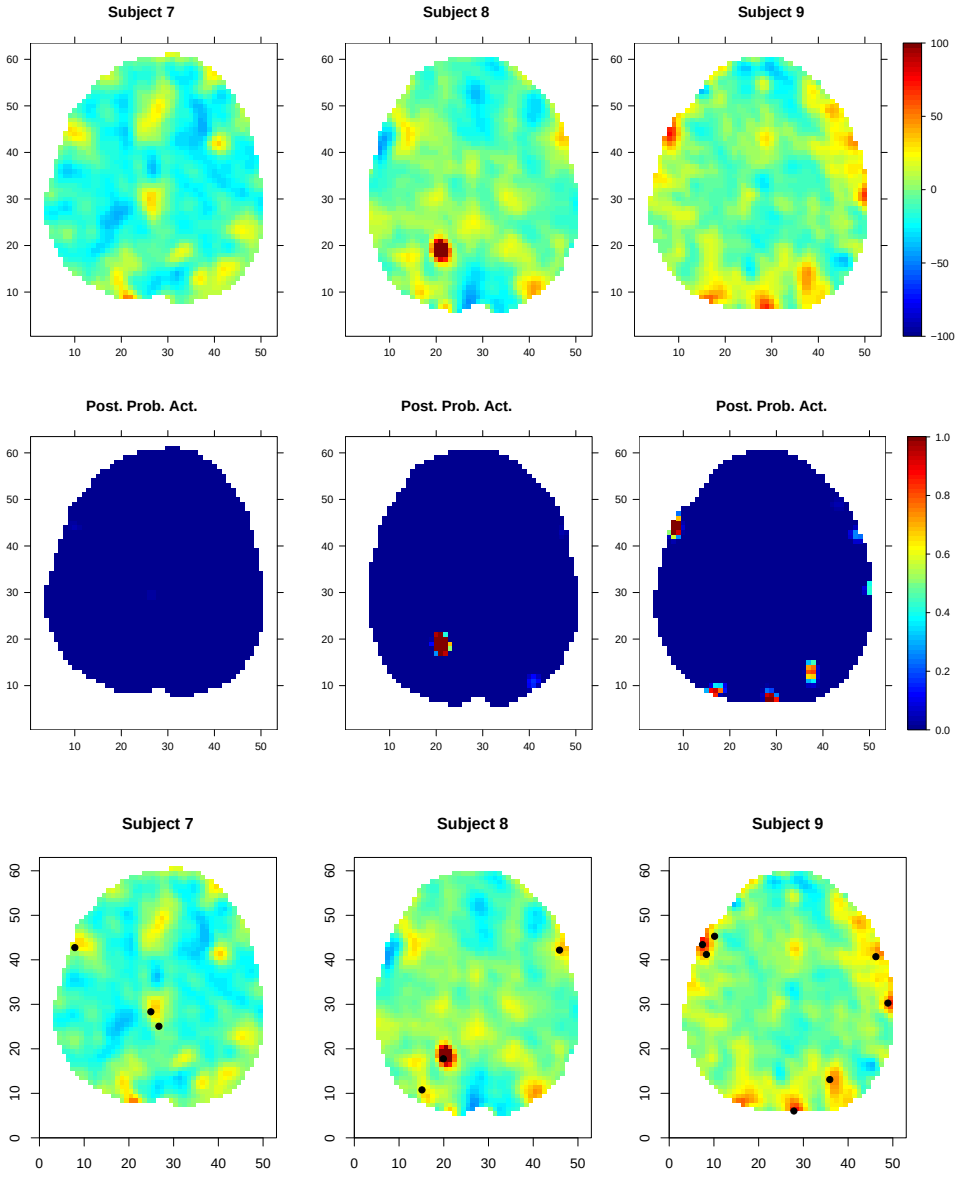


Figure D.3: Top row: The intensity data for subject 7-9. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

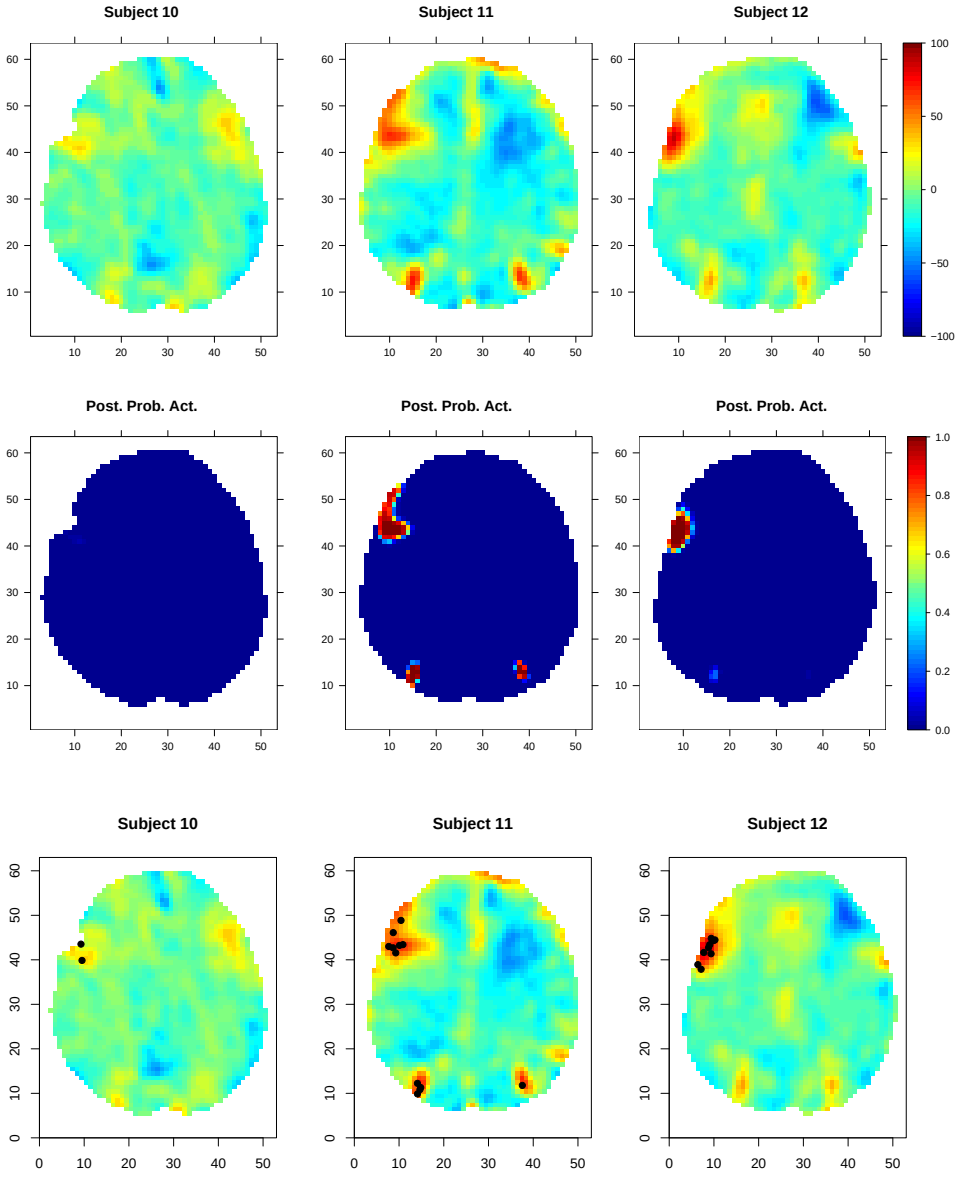


Figure D.4: Top row: The intensity data for subject 10-12. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

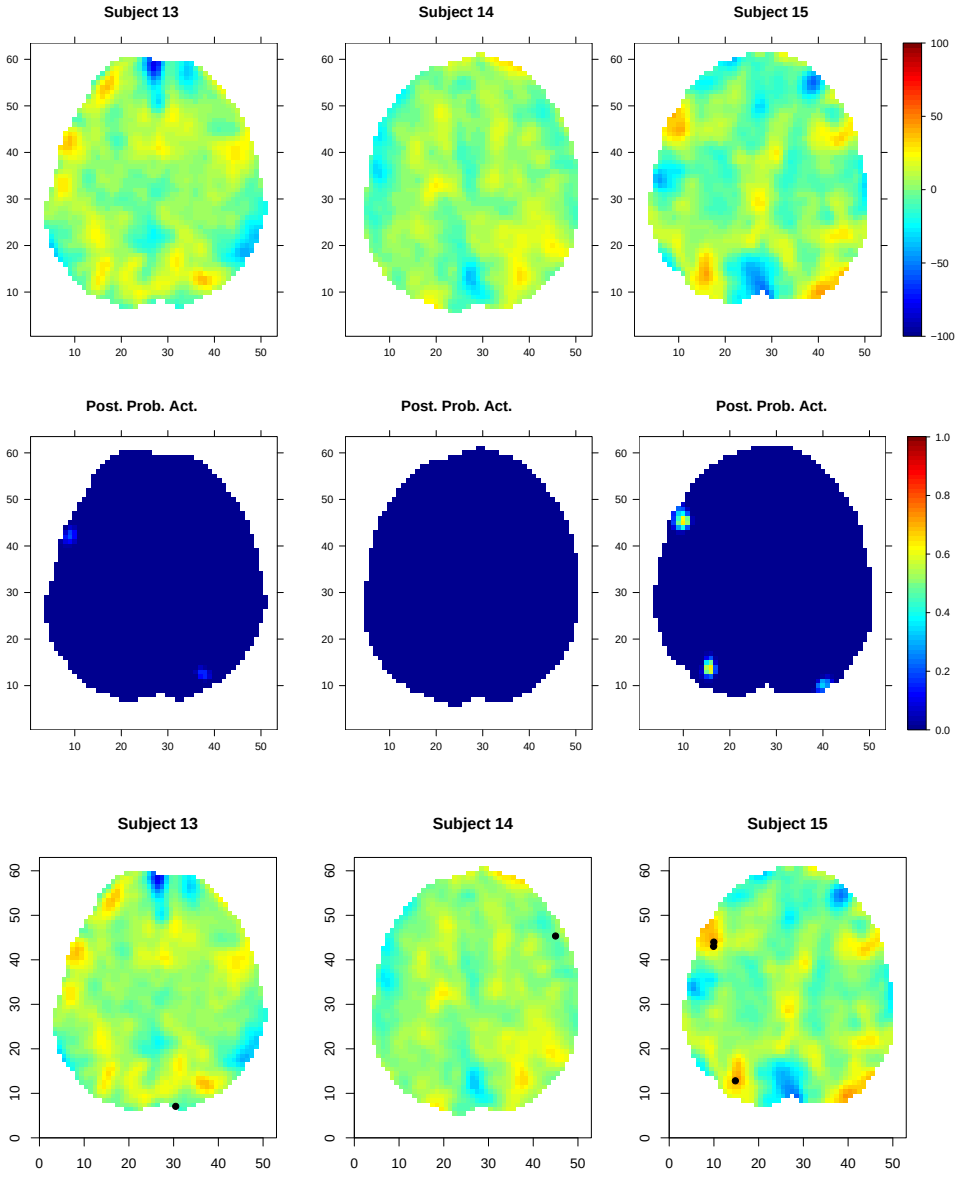


Figure D.5: Top row: The intensity data for subject 13-15. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

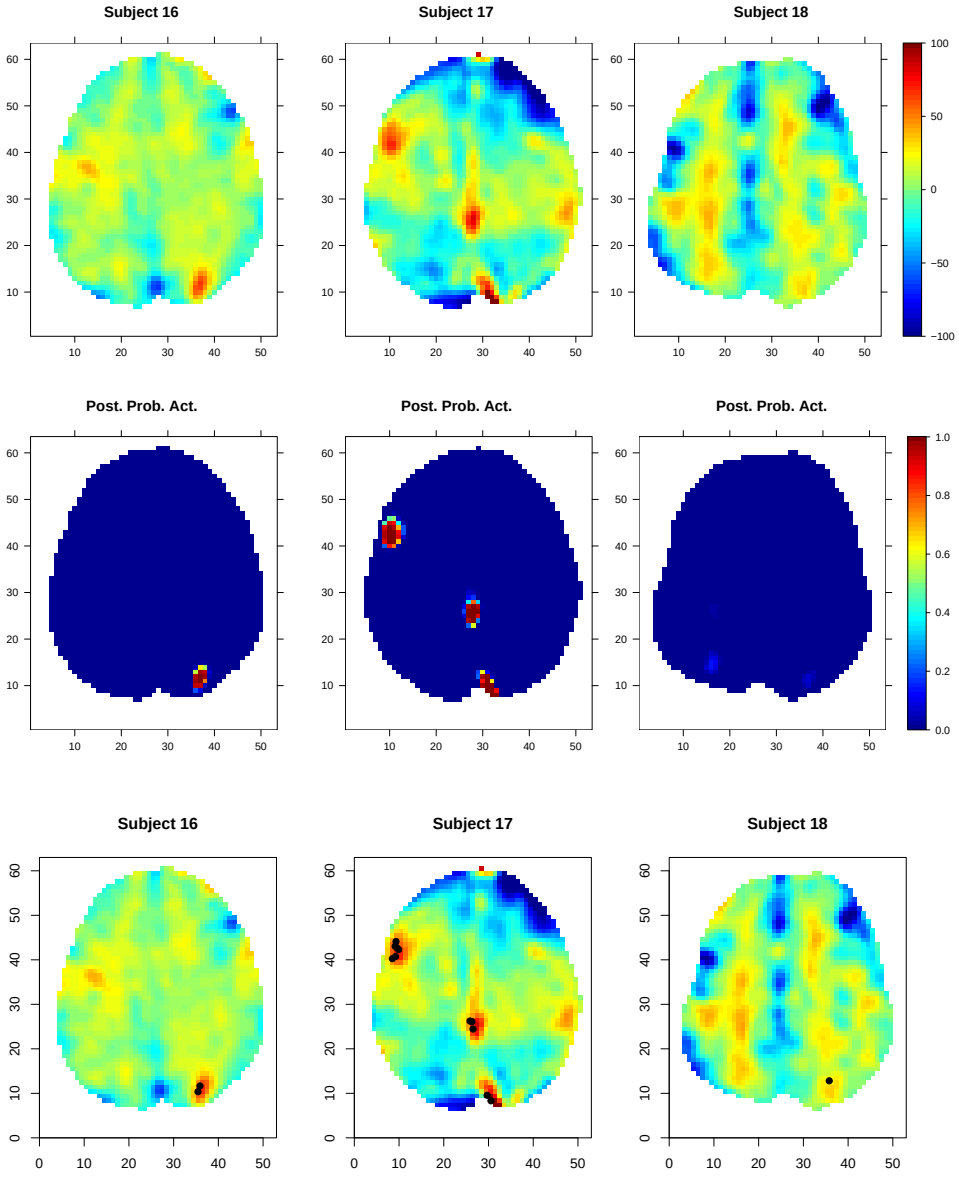


Figure D.6: Top row: The intensity data for subject 16-18. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

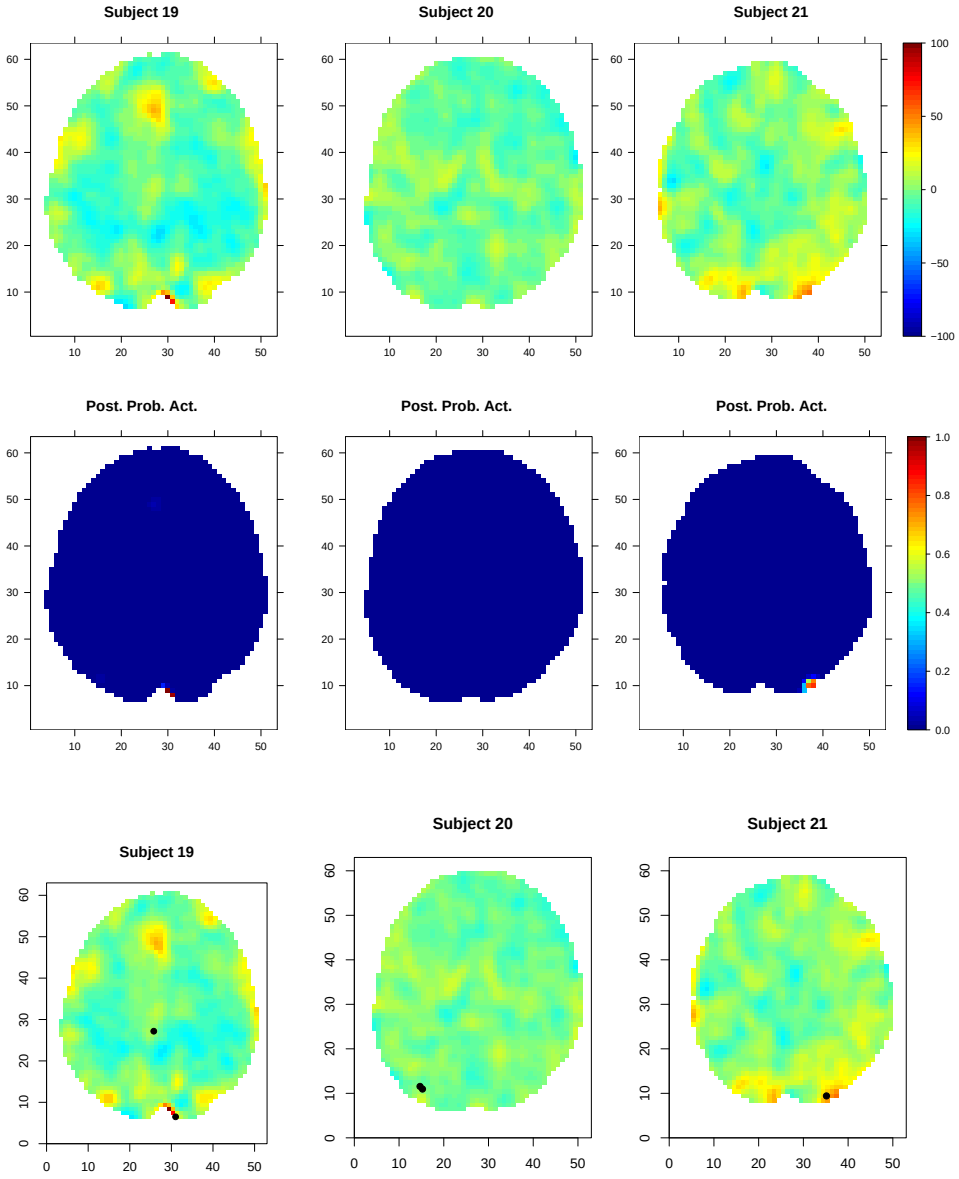


Figure D.7: Top row: The intensity data for subject 19-21. Middle row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom row: The center location of individual components at iteration 6000.

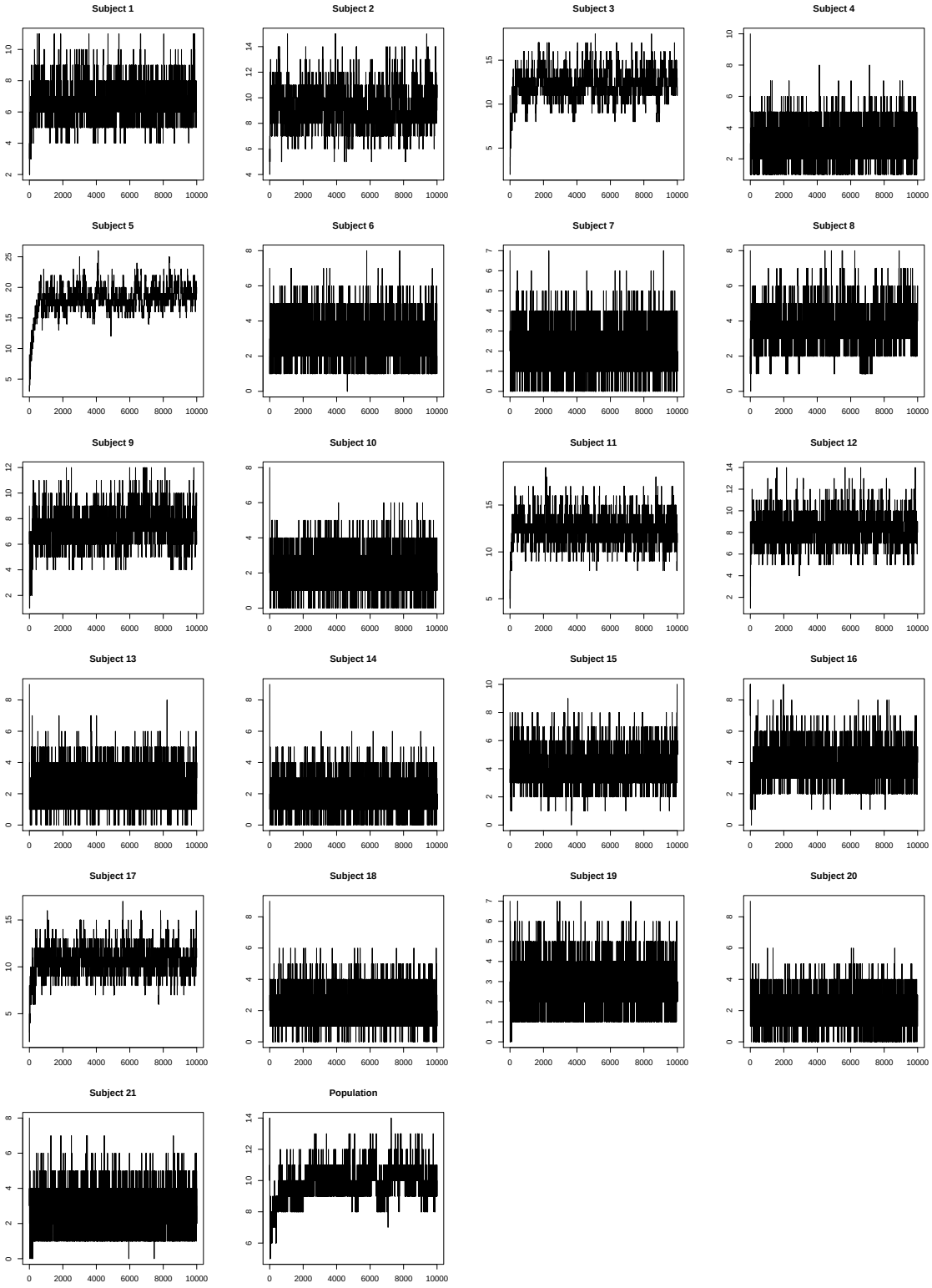


Figure D.8: Trace plot of the number of population centers and the center of individual components including the 5000 burn-in iterations

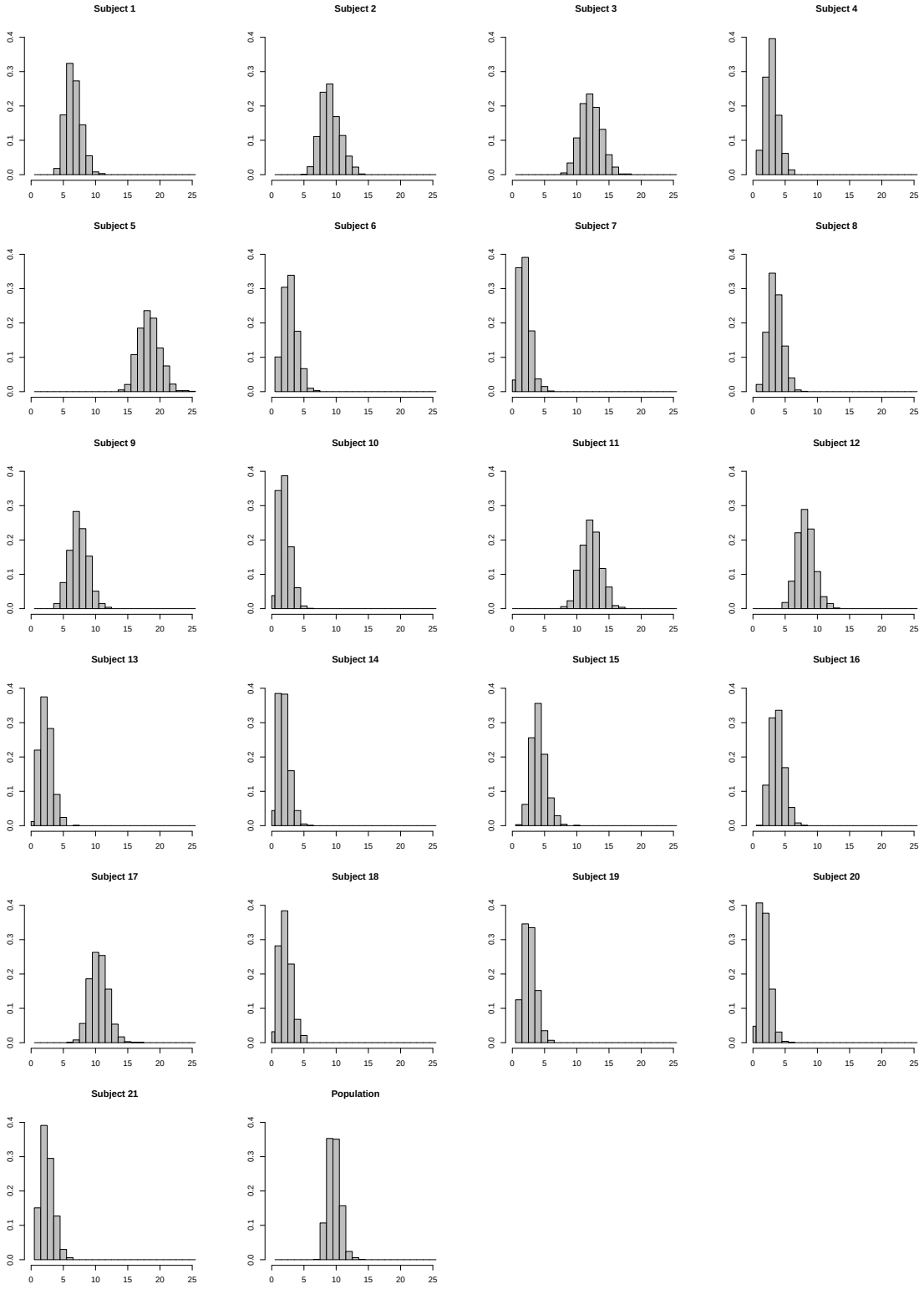


Figure D.9: Posterior distributions of the number of population centers and the center of individual components

APPENDIX E

Simulation results of all 16 subjects for the finite mixture model in Chapter II

In this Appendix we show the simulation results for the finite mixture model introduced in Chapter II. Figures E.1 to E.4 show the source data, the marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$ and the center location of individual components at iteration 7000 and iteration 7250. Figure E.5 shows the trace plots of the number of population centers and the center of individual components. Figure E.6 shows the posterior distributions of the number of population centers and the center of individual components.

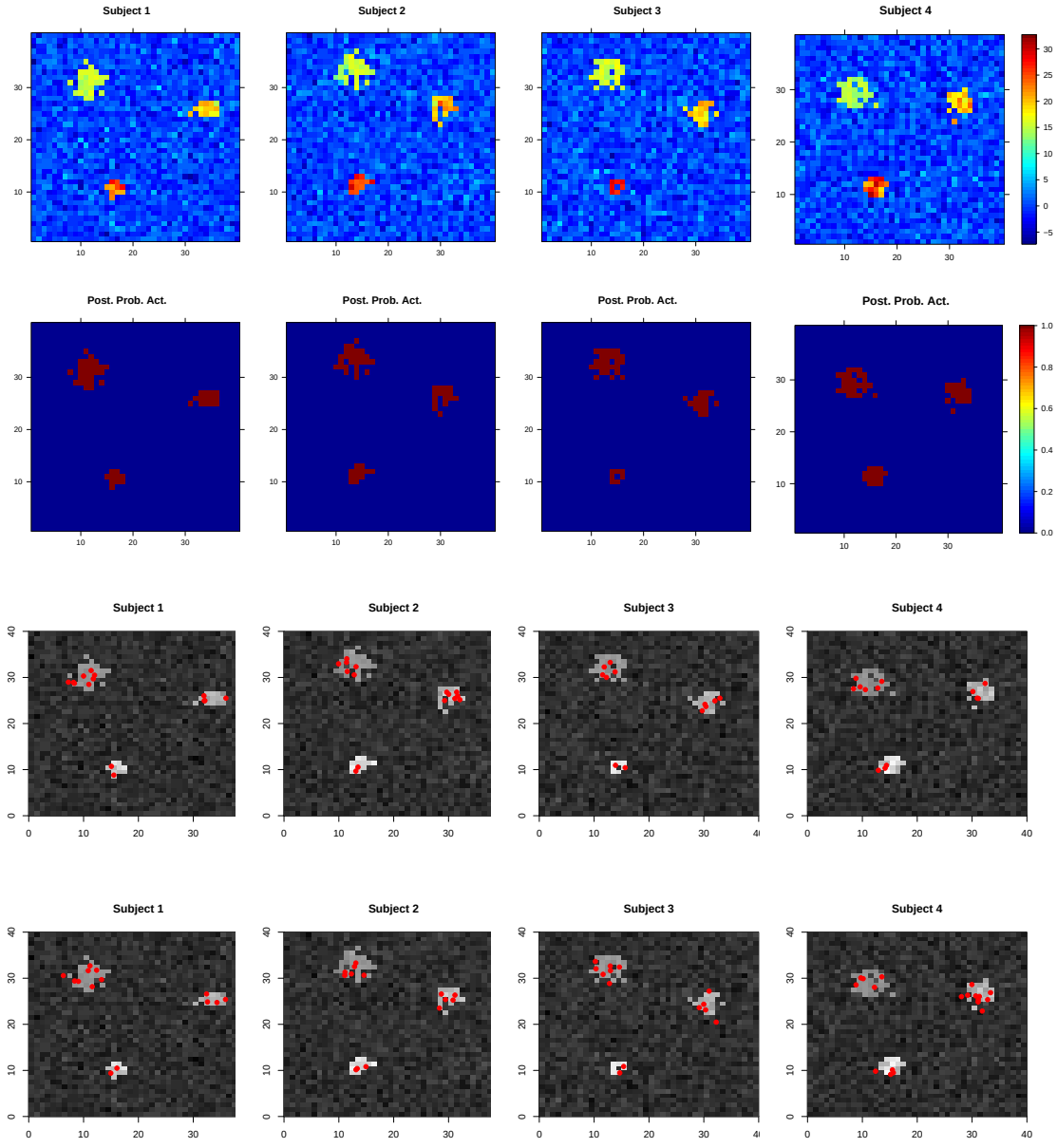


Figure E.1: First row: The intensity data for subject 1-4. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

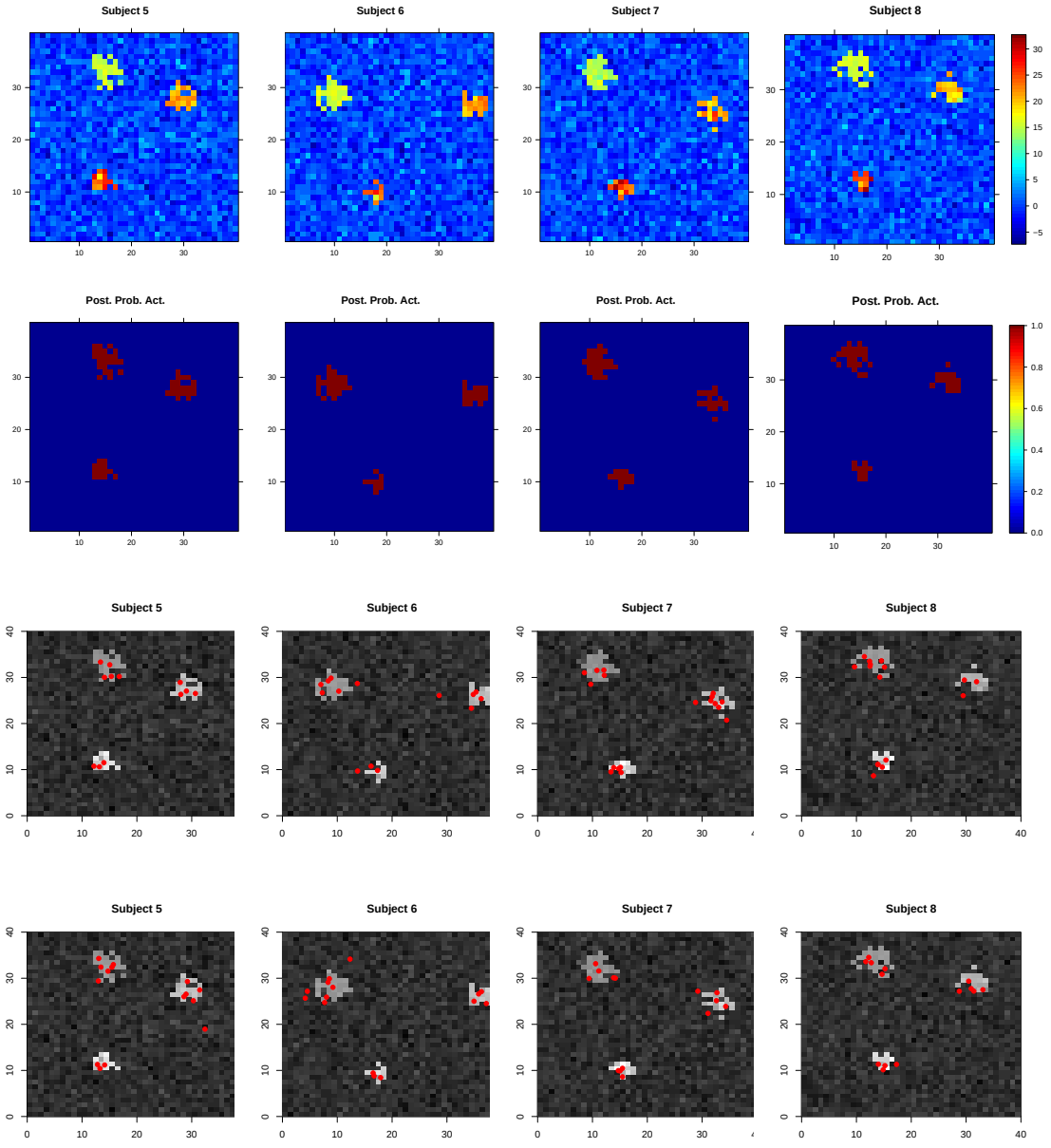


Figure E.2: First row: The intensity data for subject 5-8. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

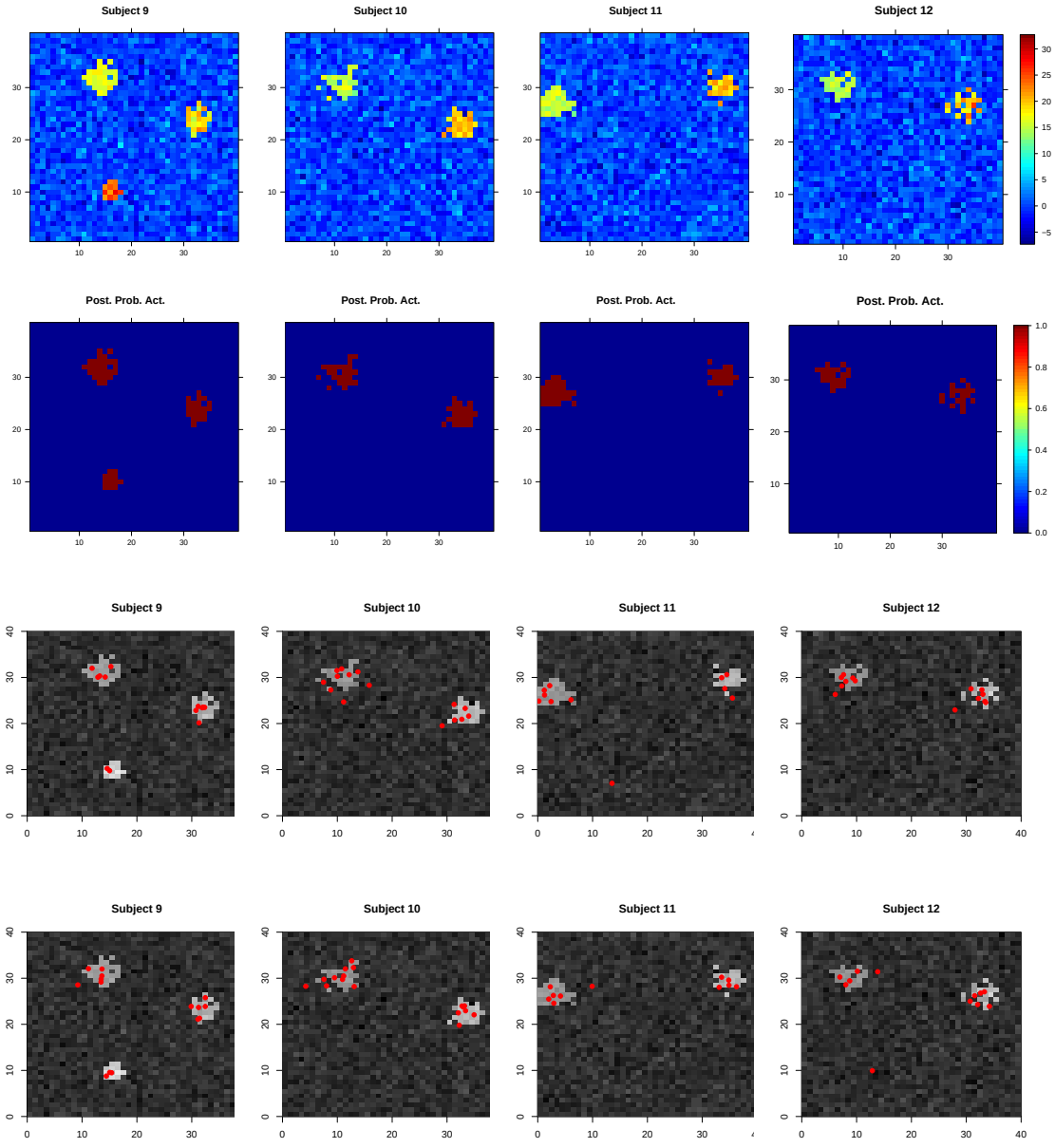


Figure E.3: First row: The intensity data for subject 9-12. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

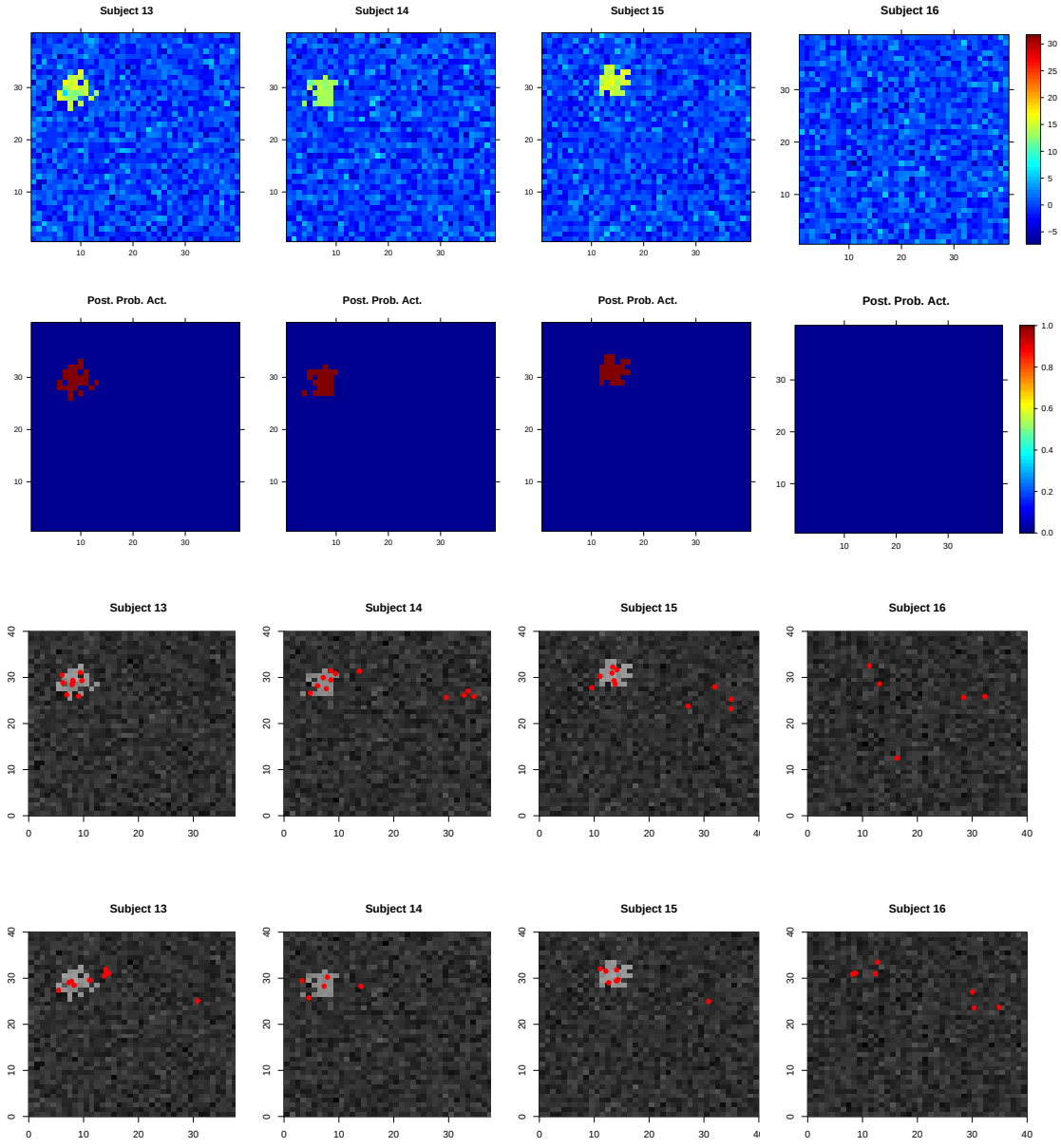


Figure E.4: First row: The intensity data for subject 13-16. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

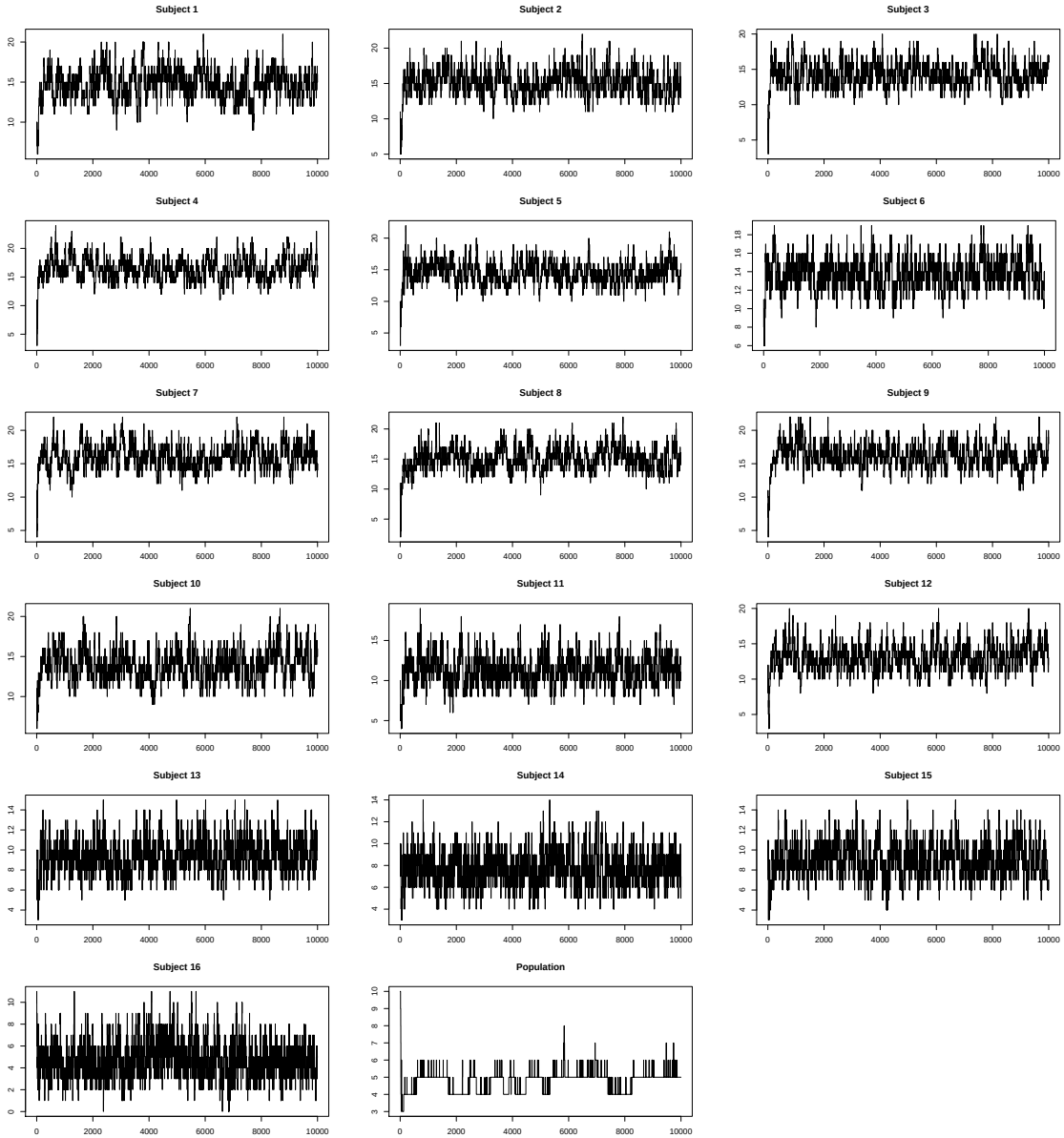


Figure E.5: Trace plot of the number of population centers and the center of Individual components including the 5000 burn-in iterations

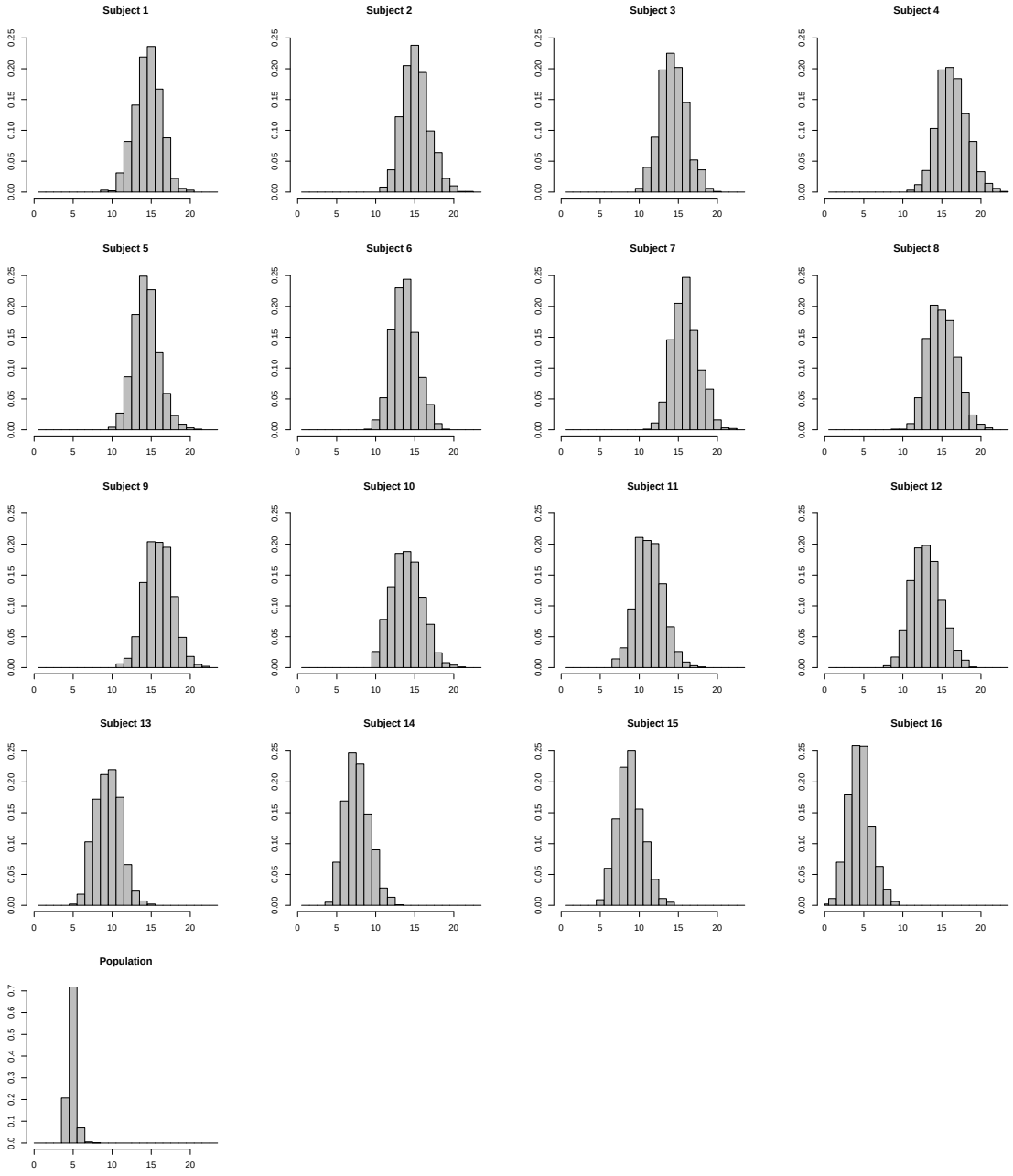


Figure E.6: Posterior distributions of the number of population centers and the center of individual components

APPENDIX F

Simulation results of all 16 subjects for the infinite mixture model in Chapter III

In this Appendix we show the simulation results for the infinite mixture model introduced in Chapter III. Figures F.1 to F.4 show the source data, the marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$ and the center location of individual components at iteration 7000 and iteration 7250. Figure F.5 shows the trace plots of the number of population centers and the center of individual components. Figure F.6 shows the posterior distributions of the number of population centers and the center of individual components.

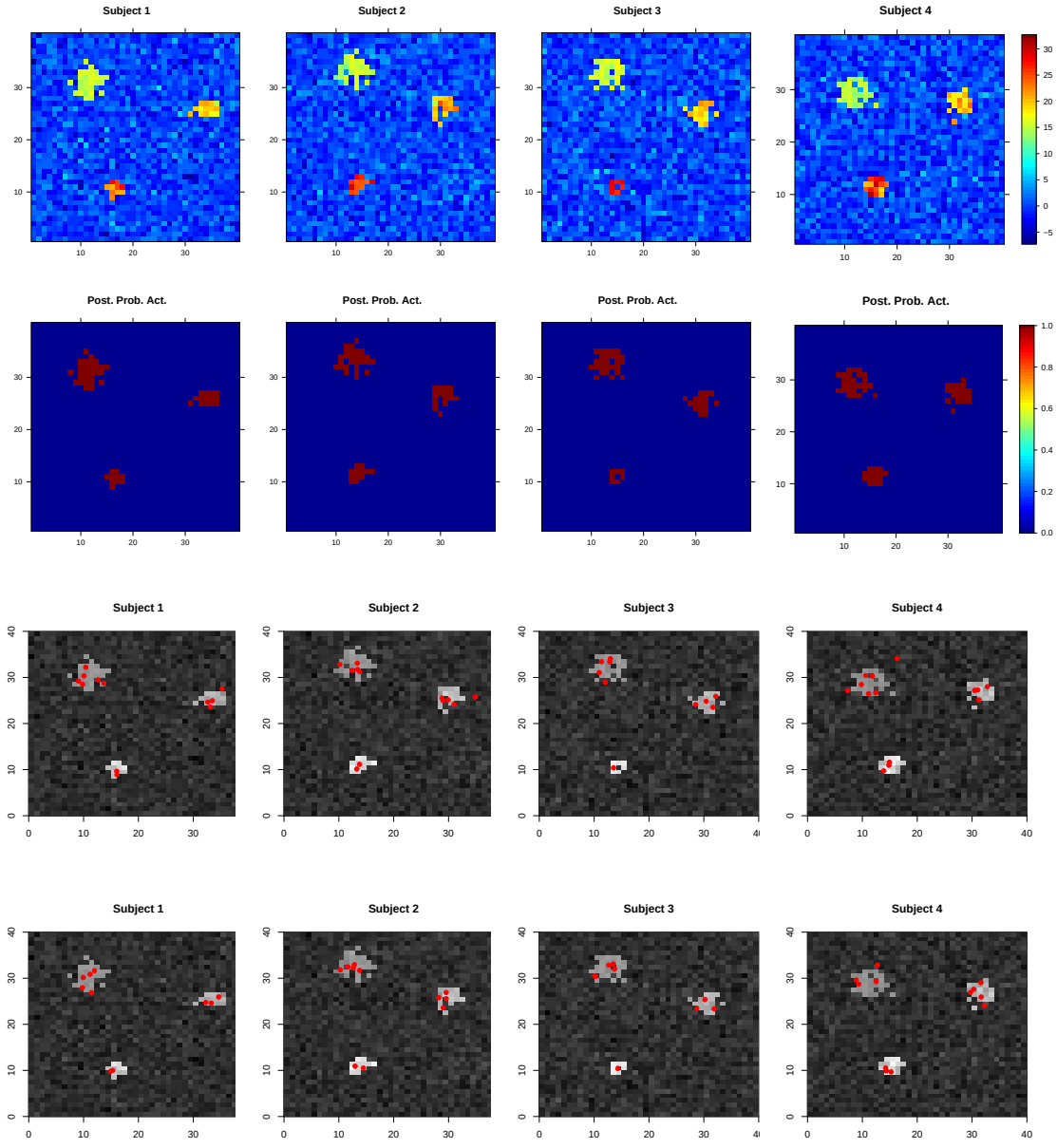


Figure F.1: First row: The intensity data for subject 1-4. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

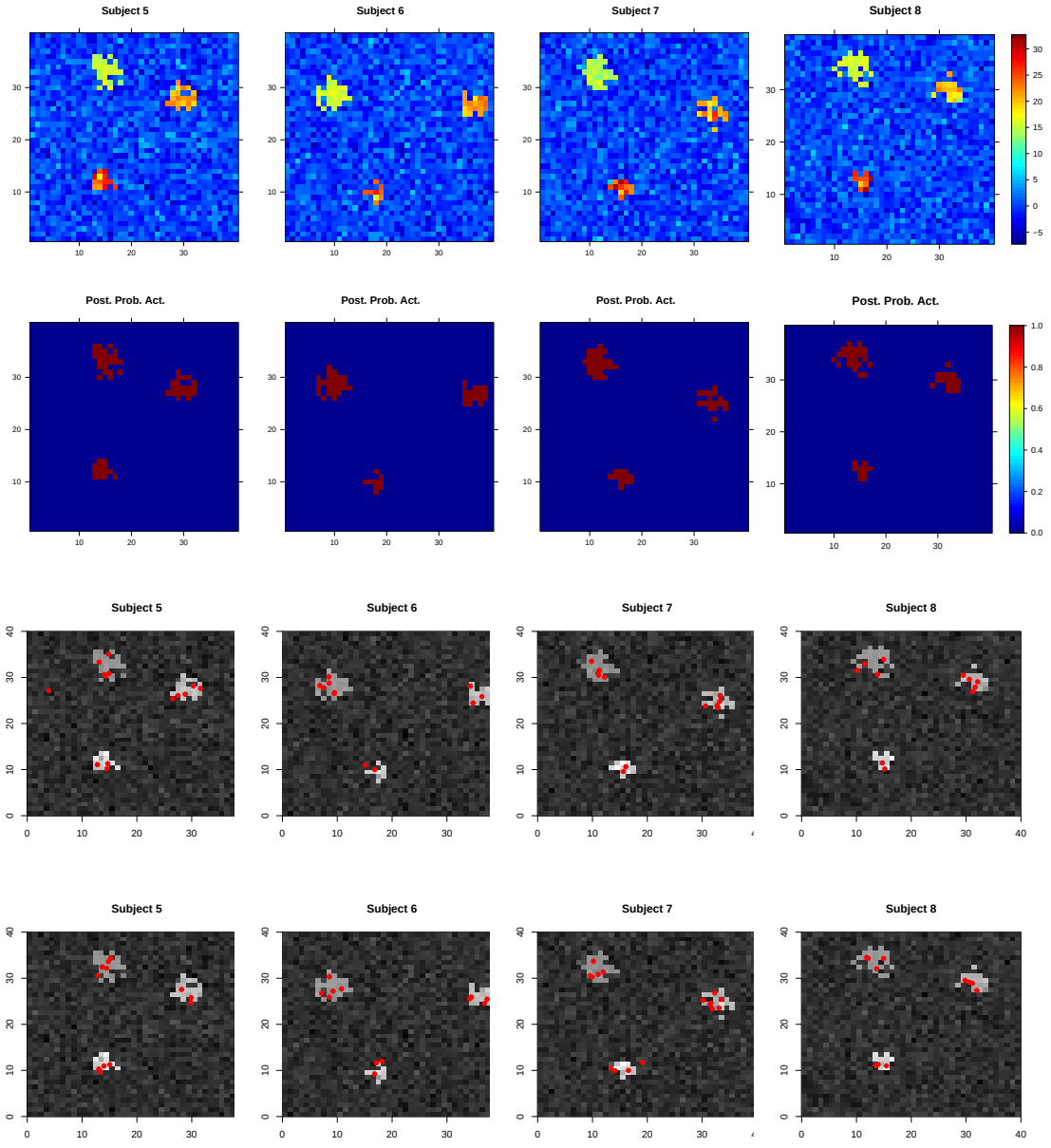


Figure F.2: First row: The intensity data for subject 5-8. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

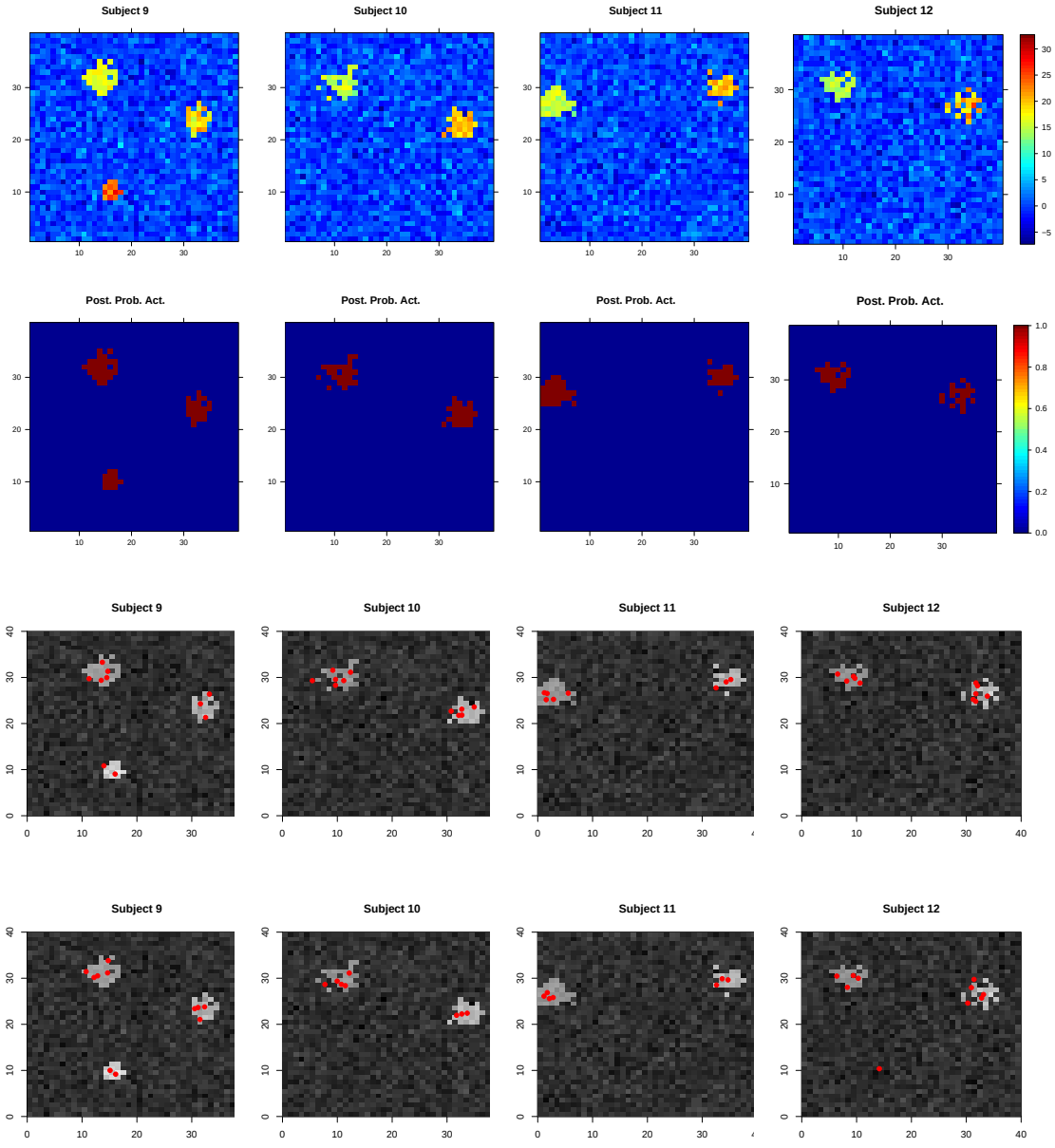


Figure F.3: First row: The intensity data for subject 9-12. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

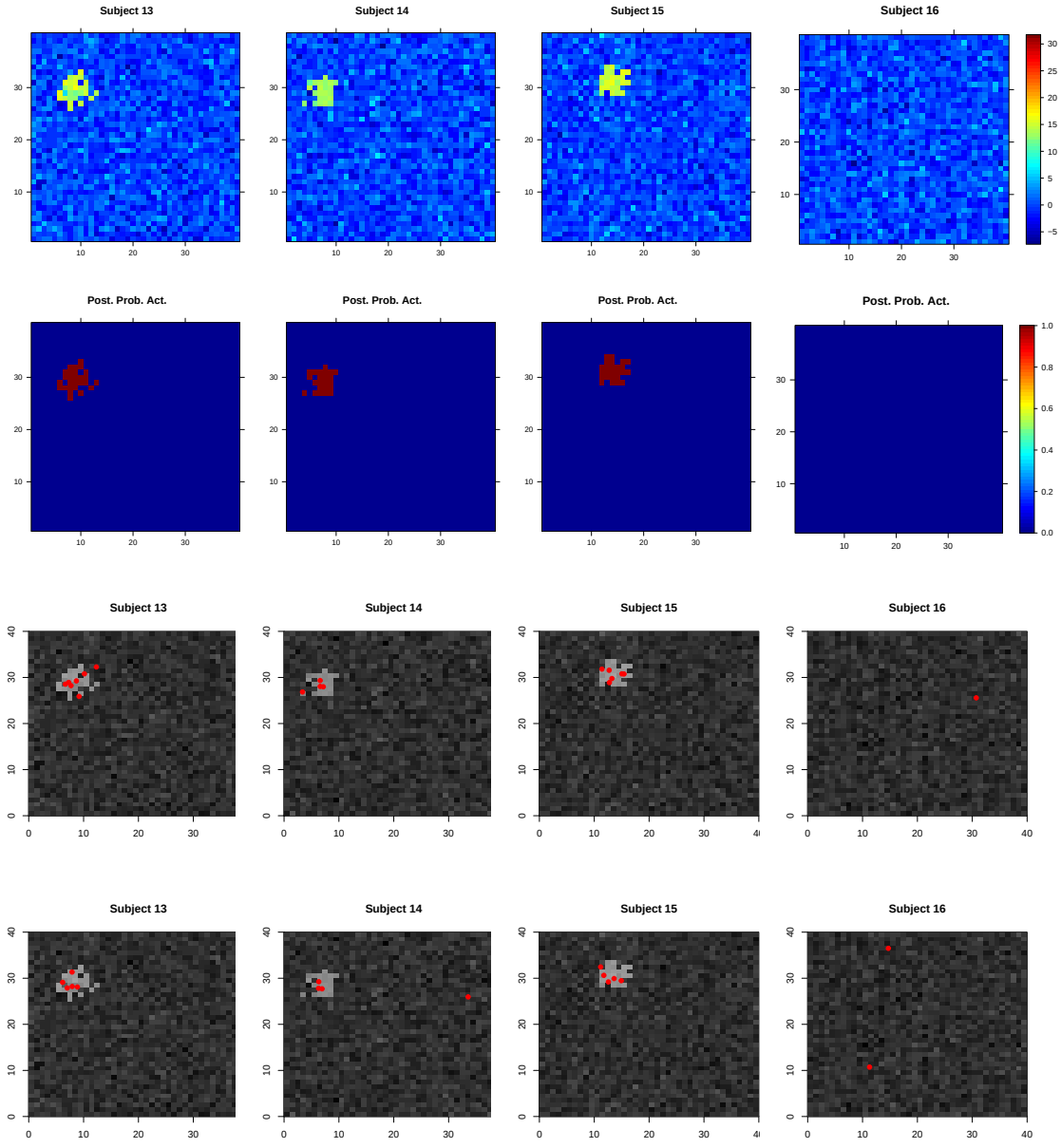


Figure F.4: First row: The intensity data for subject 13-16. Second row: The marginal posterior probability of activation: $\Pr(\omega_{jv} > 0 \mid \mathbf{y})$. Bottom two rows: The center location of individual components at iteration 7000 and at iteration 7250.

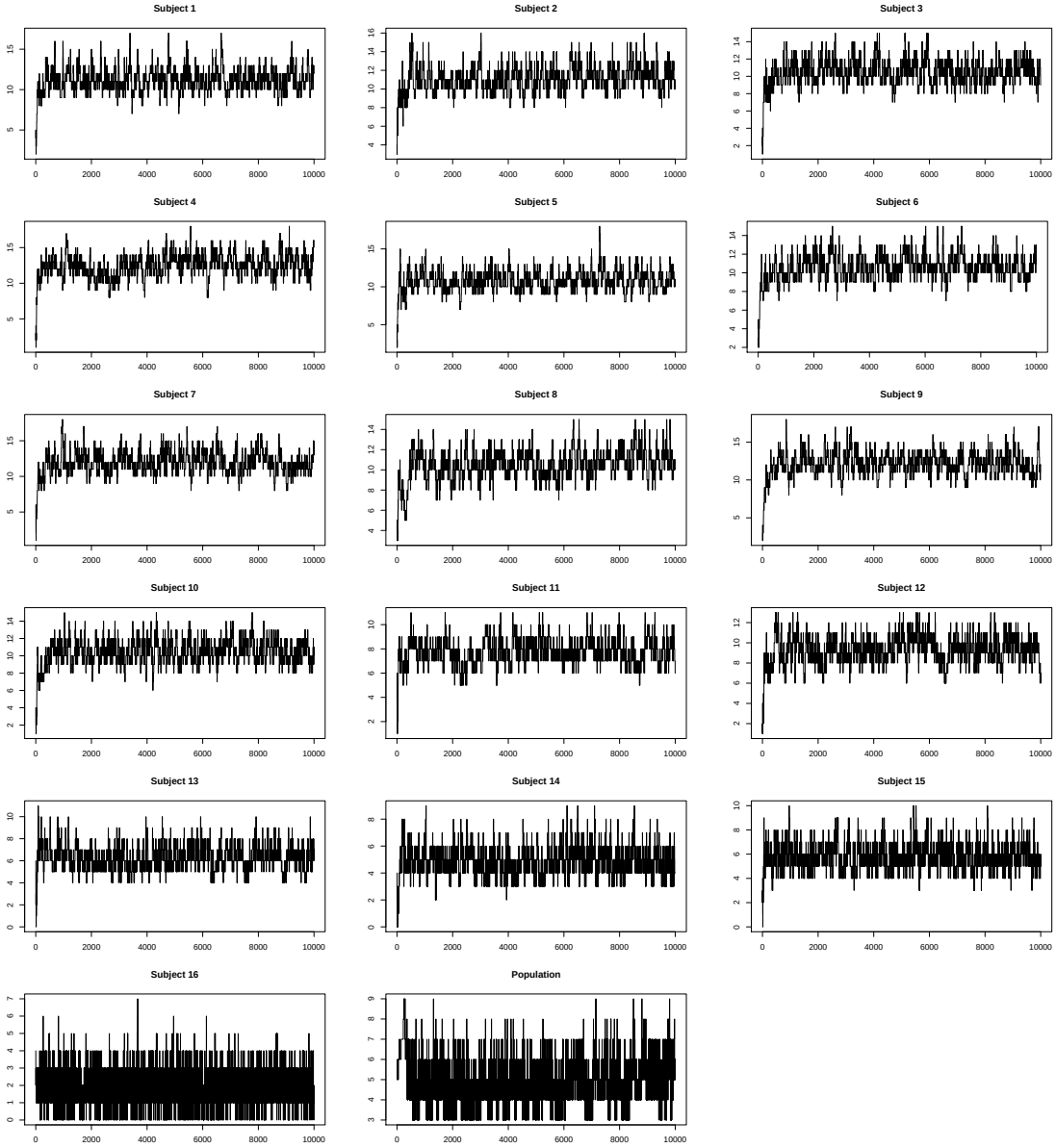


Figure F.5: Trace plot of the number of population centers and the center of individual components including the 5000 burn-in iterations

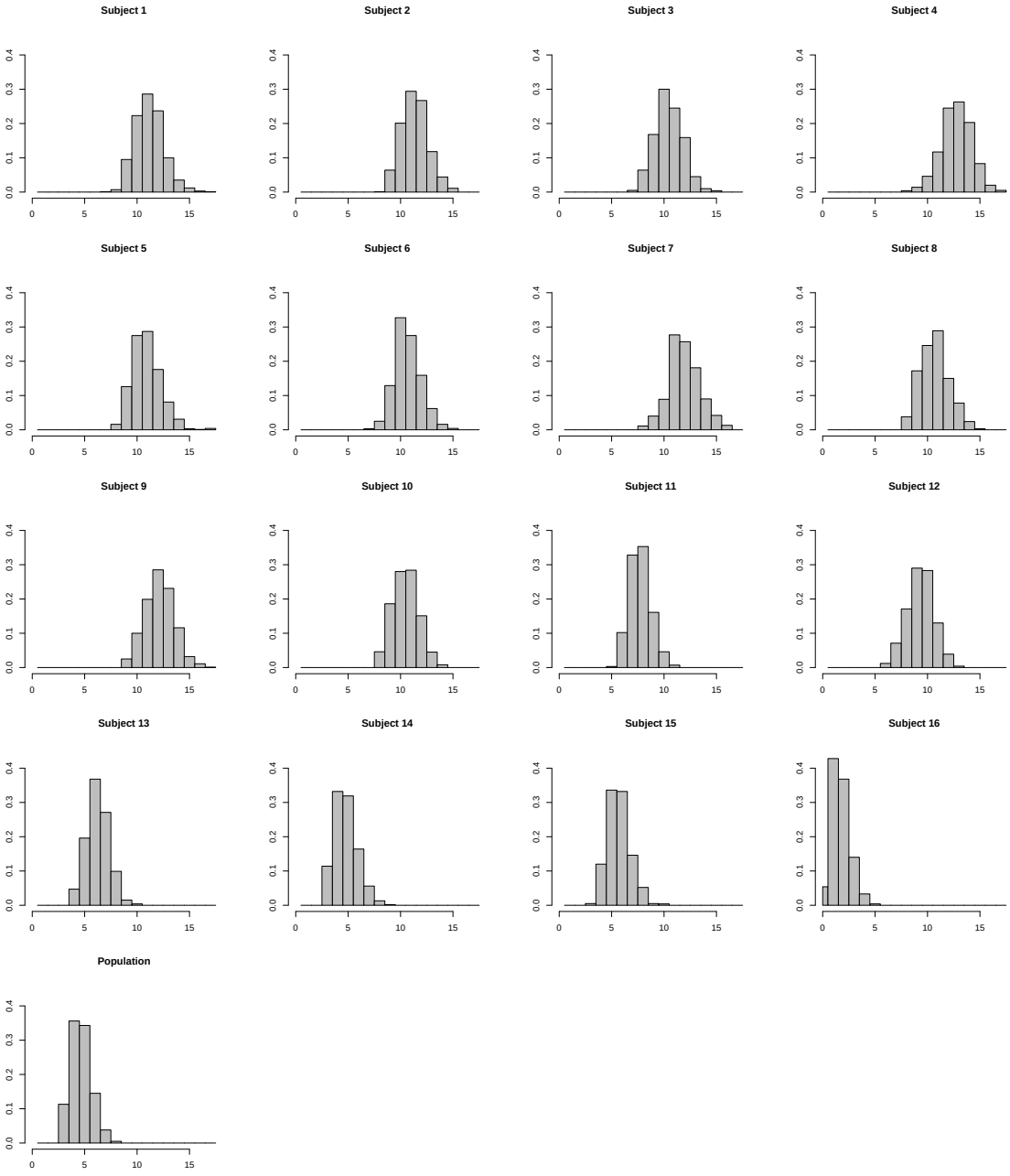


Figure F.6: Posterior distributions of the number of population centers and the center of individual components

BIBLIOGRAPHY

BIBLIOGRAPHY

[1] C. E. Antoniak. Mixtures of dirichlet processes with applications to non-parametric problems. *Annals of Statistics*, 2:1152–1174, 1974.

[2] H Attias. *A variational Bayesian framework for graphical models. Advances in Neural Information Processing Systems*. MIT Press, Cambridge, MA, 2000.

[3] A. J. Baddeley and M. N. M. van Lieshout. Stochastic geometry models in high-level vision. *Statistics and Images, Advances in Applied Statistics, a supplement to the Journal of Applied Statistics*, 20:231–256, 1993.

[4] D. Blackwell and J. B. MacQueen. Ferguson distribution via polya urn schemes. *The Annals of Statistics*, 1:353–355, 1973.

[5] E. R. Cosman, J. W. Fisher, and W. M. Wells. Exact map activity detection in fmri using a glm with an ising spatial prior. *In Proc. MICCAI'04*, 2:703–710, 2004.

[6] D.R. Cox. Some statistical models related with series of events. *Journal of the Royal Statistical Society. B*, 17:129–164, 1955.

[7] A. Dasgupta and A. E. Raftery. Detecting features in spatial point processes with clutter via model-based clustering. *Journal of the American Statistical Association*, 93:294–302, 1998.

[8] X. Descombes, F. Kruggel, and D. von Cramon. Spatio-temporal fmri analysis using markov random fields. *IEEE transactions on medical imaging*, 17:1028–1039, 1998.

[9] M. D. Escobar. Estimating normaal means with a dirichlet process prior. *Journal of the American Statistical Association*, 89:268–277, 1994.

[10] M. D. Escobar and M. West. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90:577–588, 1995.

[11] T. S. Ferguson. A bayesian analysis of some nonparametric problems. *Annals of Statistics*, 1:209–230, 1973.

[12] C. Fernandez and P. Green. Modelling spatially correlated data via mixtures:a bayesian approach. *Journal of the Royal Statistical Society. B*, 64:805–826, 2002.

[13] G. Flandin and W. D. Penny. Bayesian fmri data analysis with sparse spatial basis function priors. *NeuroImage*, 34:1108–1125, 2007.

[14] K. J. Friston, A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. R. Williams, R. S. J. Frackowiak, and R. Turner. Analysis of fmri time-series revisited. *NeuroImage*, 2:45–53, 1995.

[15] K. J. Friston, K. J. Worsley, R. S. J. Frackowlak, J. C. Mazziotta, and A. C. Evens. Assessing the significance of focal activations using their spatial extent. *Human Brain Mapping*, 1:210–220, 1993.

[16] C.R. Genovese, N. Lazar, and T.E. Nichols. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. *NeuroImage*, 15:870–878, 2002.

[17] P. J. Green. Reversible jump markov chain monte carlo computation and bayesian model determination. *Biometrika*, 82:711–732, 1995.

[18] N. V. Hartvig. A stochastic geometry model for functional magnetic resonance images. *Board of the Foundation of the Scandinavian Journal of Statistics*, 29:333–353, 2002.

[19] N. V. Hartvig and J. L. Jensen. Spatial mixture mixture modelling of fmri data. *Human Brain Mapping*, 11:233–248, 2000.

[20] W. K. Hastings. Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57:97–109, 1970.

[21] A. B. Lawson. Discussion contribution. *Journal of the Royal Statistical Society. B*, 55:61–62, 1993.

[22] S. N. MacEachern and P. Muller. Estimating mixture of dirichlet process models. *Journal of Computational and Graphical Statistics*, 7:223–238, 1998.

[23] D. L. Miglioretti, C. McCulloch, and S. L. Zeger. Combining images across multiple subjects: a study of direct cortical electrical interference. *Journal of the American Statistical Association*, 97:125–135, 2002.

[24] P. Morosan, J. Rademacher, A. Schleicher, K. Amunts, T. Schormann, and K. Zilles. Human primary auditory cortex: Cytoarchitectonic subdivisions and mapping into a spatial reference system. *NeuroImage*, 13:684–701, 2001.

[25] P. Muller, A. Erkanli, and M. West. Bayesian curve fitting using multivariate normal mixtures. *Biometrika*, 83:67–80, 1996.

[26] P. Muller and F. A. Quintana. Nonparametric bayesian data analysis. *Statistical Science*, 19:95–110, 2004.

[27] R.M. Neal. Markov chain sampling methods for dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, 9:249–265, 2002.

[28] J. Neyman and E. L. Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society. B*, 20:1–43, 1958.

[29] T.E. Nichols and Hayasaka. Controlling the familywise error rate in functional neuroimaging: A comparative review. *Statistical Methods in Medical Research*, 12:419–446, 2003.

[30] D. C. Park, R. C. Welsh, C. Marshuetz, A. H. Gutchess, J. Mikels, T. Polk, D. C. Noll, and S. F. Taylor. Working memory for complex scenes: Age differences in frontal and hippocampal activations. *Journal of Cognitive Neuroscience*, 15:1122–1134, 2003.

[31] W. D. Penny, G. Flandin, and N. J. Trujillo-Barreto. Bayesian comparison of spatially regularized general linear models. *Human Brain Mapping*, 28:275–293, 2007.

[32] W. D. Penny, N. J. Trujillo-Barreto, and K. J. Friston. Bayesian fmri time series analysis with spatial priors. *NeuroImage*, 24:350–362, 2005.

[33] J. Polzehl and V. Spokoiny. Functional and dynamic magnetic resonance imaging using vector adaptive weights smoothing. *Applied Statistics*, 50:485–501, 2001.

[34] K. Tabelow, J. Polzehl, U. Voss, H, and V. Spokoiny. Analyzing fmri experiments with structural adaptive smoothing procedures. *NeuroImage*, 33:55–62, 2006.

[35] M. N. M. van Lieshout. Stochastic geometry models in image analysis and spatial statistics. *Technical report, CWI tract 108, Amsterdam*, 1995.

[36] M. N. M. van Lieshout and A. J. Baddeley. Markov chain monte carlo methods for clustering of image features. *Proceedings of the 5th IEE international Conference on Image Processing and Its Applications*, IEE Press:241–245, 1995.

[37] M. W. Woolrich, T. E. J. Behrens, C. F. Beckmann, and S. M. Smith. Mixture models with adaptive spatial regularisation for segmentation with an application to fmri data. *IEEE transactions on medical imaging*, 24:1–11, 2005.

[38] K.J. Worsley, S. Marrett, P. Neelin, A.C. Vandal, K.J. Friston, and A.C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.