Some Novel Spatial Stochastic Models for Functional Neuroimaging Analysis

by

Jian Kang

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Biostatistics)
in The University of Michigan
2011

Doctoral Committee:

Research Associate Professor Timothy D. Johnson, Chair
Professor Peter Xuekun Song
Associate Professor Bhramar Mukherjee
Associate Professor Stephan F. Taylor
Principal Research Fellow Thomas E. Nichols, University of Warwick

To Wei, Boda and My Parents

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Dr. Timothy D. Johnson for his truly outstanding guidance on my Ph.D. study. In addition to biostatistics, I have learned a lot of useful computational techniques as well as academic writing and presentation skills from him. And I have benefited so much from his deep insights on the Bayesian thinking, his great passion for statistical computing, and his academic professionalism. Also, I wish to express my great appreciation to my mentor Dr. Thomas E. Nichols for his invaluable suggestions on my dissertation research, his overwhelming enthusiasm on imaging problems, and his broad knowledge in the neurosciences. Drs. Johnson and Nichols have introduced me into the area of neuroimaging and suggested the topics of this dissertation. I am deeply appreciative of their continuous encouragement, support and help in my Ph.D. study.

I wish to thank Drs. Bhramar Mukherjee, Peter X-K. Song, and Stephan Taylor for serving as members of my dissertation committee and providing me very helpful comments on my dissertation. Especially, I am very grateful to Dr. Mukherjee for many useful discussions on spatial point processes and Bayesian modeling, to Dr. Song for teaching me nonparametric statistics and correlated data analysis, and providing valuable suggestions on my professional development, and to Dr. Taylor for his thoughtful comments on neuroimaging meta analysis. Also, I want to thank Drs. Tor D. Wager and Lisa Feldman Barrett for their expertise in neuropsychology, insights on emotion analysis, and stimulating suggestions on my research.

My appreciation also goes to the Department of Biostatistics at the University of Michigan which has provided a strongly supportive environment for study and research. I have had an intellectually stimulating, meaningful, and enjoyable experience as a graduate student in this department. I wish to thank all of the knowledgable faculty members for their expertise in biostatistics. In particular, I have had many enlightening conversations with Dr. Yun Li on the Bayesian identifiability issues, with Drs. Lu Wang and Wen Ye on longitudinal and functional data analysis. I am also grateful to all the staff members, fellow students and friends in the department for their sincere help and support.

Also, I would like to thank my brother and my parents for their love and consistent encouragement. My parents has made every effort to offer me the best educational opportunities. My brother Boda has taught me various mathematical skills and brought me into the world of statistics. Last but not least, I am greatly indebted to my wife Wei for her boundless love and continuous support of my learning and research activities during these years.

TABLE OF CONTENTS

DEDICATION . . . . **ii**

ACKNOWLEDGEMENTS . . . . **iii**

LIST OF FIGURES . . . . **vii**

LIST OF TABLES . . . . **x**

LIST OF APPENDICES . . . . **xi**

CHAPTER

I. Introduction 1

1.1 Functional Neuroimaging 2
1.2 Statistical Parametric Mapping 5
1.3 Meta Analysis of Functional Neuroimaging Data 6
1.4 Spatial Point Processes 8
1.5 Dissertation Outline 10

II. Meta Analysis of Functional Neuroimaging Data via Bayesian Spatial Point Processes 15

2.1 Introduction 16
2.2 The Model 20
  2.2.1 Model Outline and Notation 20
  2.2.2 Model Details 22
2.3 Application 25
  2.3.1 Prior Parameters 26
  2.3.2 Analysis of the Emotion Meta Analysis Dataset 28
  2.3.3 Model Assessment 33
2.4 Simulation Studies And Sensitivity Analysis 34
2.5 Discussion 35

III. Hierarchical Poisson/Gamma Random Field Models For Multi-type Spatial Point Patterns 44

3.1 Introduction 44
3.2 The Model 48
  3.2.1 Hierarchical Poisson/Gamma Random Fields 48
  3.2.2 HPGRF for multiple independent realizations 57
3.3 Examples 59
  3.3.1 Bramble Canes 59
  3.3.2 Functional Neuroimaging Meta Analysis of Emotion 61

3.4 Simulation Studies 63
    3.4.1 Simulated Data Sets 63
    3.4.2 Accuracy of Intensity Estimates 64
3.5 Discussion 64

IV. A Bayesian Spatial Point Process Classifier with Application to Functional Neuroimaging Reverse Inference 73

4.1 Introduction 74
4.2 The Model 78
    4.2.1 Bayesian Spatial Point Process Classifier 79
    4.2.2 The BHICP Model 83
    4.2.3 The HPGRF Model 85
4.3 Single Focus Impact Map 85
4.4 Application 87
4.5 Simulation Studies 88
    4.5.1 Simulated Data Sets 88
    4.5.2 Simulation Performance 90
4.6 Discussion 91

V. Conclusion 98

APPENDICES **102**

BIBLIOGRAPHY **145**

LIST OF FIGURES

Figure

1.1 An illustration of the BOLD fMRI mechanism. (A): BOLD signal mechanism in MRI; (B): oxygenated blood (red) and deoxygenated blood (blue) flow during rest and activation. 11

1.2 An illustration of fMRI signal and single subject data. 12

1.3 The fMRI data processing pipeline illustrates the paths from the raw fMRI time series to the statistical parametric maps to the peak activation locations (foci) via preprocessing, statistical analysis and cluster analysis. 13

1.4 Functional neuroimaging meta analysis data: foci reported from different studies. . 14

2.1 Panel (A): A subset of the emotion meta analysis data set. Panel (B): All foci, (x, y, z) locations, from all studies plotted in the MNI brain template. 39

2.2 Hierarchical Model Illustration: Level 1: The foci from each study, $\mathbf{x}_c$, are the observed data, here shown as open and filled circles. The open circles represent singly reported foci, $\mathbf{x}_c^0$, and the solid circles represent multiply reported foci, $\mathbf{x}_c^1$. Whether foci are multiply or single reported is a latent property, δ_x. Level 2: Multiple reported foci cluster about latent study activation centers, $\mathbf{y}_c$, (open triangles) with the dashed circles representing $\mathbf{\Psi}_y$. Singly reported foci in level 1 are shown in level 2 as open circles as they, along with the study activation centers, may cluster together in level 3. Level 3: Activation centers, $\mathbf{x}_c^0$ from level 1 and $\mathbf{y}_c$ from level 2, may cluster about a population center, $\mathbf{z}$ (filled diamond) with the dashed circle representing $\mathbf{\Sigma}_z$ or may fail to cluster and are modeled as background scatter and outliers. 40

2.3 Qualitative comparison of the modALE map, MKDA map, the posterior expected activation center intensity function, the posterior predictive intensity for a new study and the posterior expected population center intensity function. We stress here that the gray scale values of the modALE, MKDA and intensity maps are not comparable as their interpretations are not comparable. Qualitatively, the first three rows are similar. The population intensities, however, are much more focused than the activation center intensities, especially in slices $Z = -22, 18, 28$. This reflects the larger variability of activation centers about the population center than the variability of the population centers, themselves. 41

2.4 The 95% marginal credible ellipses. Large, ellipses are the marginal ellipses of the activation centers. Small, ellipses are the marginal ellipses of the population centers within each amygdala. The black regions (masks), covered by the white ellipses, are the amygdalae. 42

2.5 The 95% marginal credible ellipses for population centers from positive (dashed ellipses) and negative (solid ellipses) emotion studies. The “x” and the circle represent the centers of the respective ellipses. 43

3.1 Locations of newly emergent and one year old bramble canes. 67

3.2 The estimate of the multivariate J function, e.g. $J_{01}(r)$ (solid line) for the bramble canes data with the 95% confidence envelope (dashed lines) based on 100 random torus translations of the data. 68

3.3 The HPGRF and PGRF posterior mean intensity estimates for each type of bramble canes point patterns. The HPGRF also provide a “population mean” intensity estimate. 68

3.4 The 1393 foci reported from 219 studies of five emotions 69

3.5 HPGRF: Foci intensity estimation for five emotions (Sad, Happy, Anger, Fear and Disgust) compared with the “population mean” intensity estimation 70

3.6 The HPGRF posterior mean intensity estimates of the two axial slices (Z = −20mm, 26mm) for five emotions (Sad, Happy, Anger, Fear and Disgust) compared with the IPGRF intensity estimates. 71

3.7 The true intensity compared with the HPGRF and IPGRF intensity estimates for the simulated Poisson cluster point process data sets. 72

4.1 Two axial slices (Z = -20mm, 26mm) of the predictive intensity of study foci, the posterior intensity of population centers, and single focus impact maps produced by BHICP model for the meta analysis of five emotions. 96

4.2 Simulated Data. 97

A.1 Sensitivity analysis results: comparisons of the posterior activation center intensity on 11 slices of the brain from $Z = -42$mm to $Z = 58$mm. 120

A.2 Sensitivity analysis results: comparisons of the population center intensity on 11 slices of the brain from $Z = -42$mm to $Z = 58$mm. 121

A.3 Sensitivity analysis results: comparisons of the 95% marginal credible ellipses of both population centers (blue circles) and activation centers (yellow circles). 122

A.4 Simulation A results: the images in the first column show the simulated foci (white points) and the true locations of population centers (red X), projected unto the XY, XZ and YZ plane for the top, middle and bottom, respectively. The images in the middle column show the posterior intensity of the activation centers, integrated over the Z, Y and X directions for the top, middle and bottom, respectively. The images in the last column show the posterior intensity of the population centers, each integrated over one dimension as in the middle column. 125

A.5 Simulation B results: The first column shows the simulated foci and the true locations of population centers. The middle column show the posterior intensity of the activation centers, and the last column shows the posterior intensity of the population centers. See Figure 4 for display conventions. 127

A.6 Simulation C results. The first column shows the simulated foci. The middle column shows the posterior intensity of activation centers, and the last columns shows the posterior intensity of the "population centers". See Figure 4 for display conventions. 128

A.7 The posterior standard deviation of the activation center intensity on 11 axial slices of the brain (from $Z = -42$mm to $Z = 58$mm). 130

LIST OF TABLES

Table

2.1 Data Summaries. 37

2.2 The 95% credible ellipsoid volumes for population and activation centers. 38

3.1 Summary statistics of the posterior distribution of the smoothing kernel variance σ_j^2 via HPGRF and PGRF for bramble canes data. 66

3.2 Number of studies and foci in the emotion emotion analysis data set. 66

3.3 Parameters for true intensity functions 66

3.4 Summary statistics of the posterior distribution of the smoothing kernel variance σ_j^2 via HPGRF and PGRF for five emotion meta analysis 67

3.5 The IMSE and IWMSE of the intensity estimation via HPGRF and IPGRF 67

4.1 Summary statistics from the emotion meta analysis data set 92

4.2 The LOOCV classification rates 93

4.3 Simulation study results: Case 1 94

4.4 Simulation study results: Case 2 95

A.1 Sensitivity Analysis Priors. 118

A.2 The summary statistics on the posterior number of population centers 119

A.3 The estimated location (x, y, z), the Euclidean distance in locations between scenario 1 and other scenarios (Dist.), the volume of 95% credible ellipsoid and the volume 95% credible interval of population center as well as activation centers conditional on the population centers located in the amygdalae. 123

A.4 The posterior expected number of population centers in different ROIs and the posterior probability that at least one population centers is located in the ROI. . . 129

LIST OF APPENDICES

Appendix

A. The Details of Bayesian Hierarchical Independent Cluster Process Model 103

- A.1 Probability Equivalence Assumption 103
- A.2 Algorithm Details 105
  - A.2.1 Posterior Distribution 105
  - A.2.2 Imputation of Missing Type Indicators 106
  - A.2.3 Spatial Birth-and-Death Processes 106
  - A.2.4 Standard MCMC Updates 111
  - A.2.5 Normal and T Probability Computation 114
  - A.2.6 Posterior Intensity Estimation 114
  - A.2.7 Credible Ellipsoid Computation 115
  - A.2.8 Pseudo Code 116
- A.3 Simulation Studies 117
  - A.3.1 Sensitivity to Priors 117
  - A.3.2 Sensitivity to Model Specification 124
- A.4 Whole Brain Results 129

B. The Details of Hierarchical Poisson/Gamma Random Fields Model 131

- B.1 Proofs 131
  - B.1.1 Proof of Theorem 1 133
  - B.1.2 Proof of Theorem 3 135
  - B.1.3 Proof of Theorem 4 136
  - B.1.4 Proof of Theorem 5 136
- B.2 PGRF for multiple spatial point patterns 137

C. The Details of Reverse Inference 141

- C.1 Proofs 141
  - C.1.1 Proof of Proposition 1 141
  - C.1.2 Proof of Corollary 1 142
  - C.1.3 Proof of Proposition 2 142
  - C.1.4 Proof of Proposition 3 143

CHAPTER I

Introduction

The desire to understand the human brain has been one of the main interests of scientists throughout the 20th century. In the last 25 years several functional neuroimaging techniques have been developed to measure the relationship between activity in certain brain regions and specific mental functions (Jezzard et al. 2001, Raichle 2006). The research of functional neuroimaging draws together a multi-disciplinary community of neuroscientists, psychologists, engineers, physicists, statisticians and computer scientists. Statisticians, without exception, have had a key role to play in this research and have already made many important contributions to the fields. A number of statistical methods have been proposed to analyze functional neuroimaging data over the last two decades, see, for example, Friston et al. (2007), Lazar (2008) and Lindquist (2008) for an overview. However, the statistical results of most neuroimaging studies are underpowered and suffer from poor reproducibility due to small sample size (Wager et al. 2007a, Yarkoni 2009). Thus, there is a growing interest in synthesis and integration of results across neuroimaging studies via statistical meta analysis (Wager et al. 2009, Salimi-Khorshidi et al. 2009, Yarkoni et al. 2010). Neuroimaging studies typically report three-dimensional coordinates for brain activation locations, i.e. points in a three-dimensional brain. The number and locations of the points are both random quantities, Thus, a natural model for such data is a spatial point process model (Møller and Waagepetersen 2004; 2007). Motivated by this problem, we develop some new spatial point process models and discuss their applications to the meta analysis of functional neuroimaging data.

The remainder of this chapter is organized as follows: we start in Section 1.1 with an overview of functional neuroimaging techniques and data acquisition procedures. The standard approaches to analyzing functional neuroimaging data in single studies are summarized in Section 1.2. The rationale and challenges of neuroimaging meta analyses are presented in Section 1.3. We provide a brief introduction to spatial point processes in Section 1.4. Finally, we outline the dissertation in Section 1.5.

1.1 Functional Neuroimaging

Functional neuroimaging has experienced a rapid growth since 1990. Currently there exist several imaging modalities that measure physiological changes due to brain activation that provide different perspectives on brain function. These modalities include functional magnetic resonance imaging (fMRI), single photon emission computed tomography (SPECT), positron emission tomography (PET), electroencephalography (EEG) and magnetoencephalography (MEG). Both EEG and MEG are based on the electrical and magnetic activity in the brain. They have excellent temporal resolution but poor spatial localization compared with PET and fMRI. These last two imaging modalities have played a prominent role in the study of human brain function in health and disease. Both PET and fMRI are able to measure brain activity indirectly by measuring aspects of blood flow, volume and oxygen content. Specifically, when a cluster of neurons fire more rapidly, both blood flow and the oxygen content of the blood in that region increase. PET uses radio-labeled tracers to identify the blood flow changes related to neural activity, while fMRI measures the changes in oxygen content of the blood. Since fMRI is a non-invasive technique that does not require the injection of radioactive tracers, it has taken a dominant position in the field of functional neuroimaging. Next, we focuses on a very short overview of how fMRI data are acquired; more detailed explanations can be found in Jezzard et al. (2001).

We briefly begin with the science of magnetic resonance imaging (MRI). Hydrogen (^{1}H), the atom with a single proton and no neutrons, is found in abundance in the body and is very sensitive to magnetic fields (Wills and Hector 1924). This makes hydrogen particularly easy to image using magnetic resonance (MR). A MR scanner produces a very powerful magnetic field that is many times greater than the natural magnetic field of the earth. When the body is placed into a MR scanner, the magnetic moments of hydrogen change and align with the direction of the field (Hashemi et al. 2004). A radio frequency (RF) pulse transmitter is turned on to inject additional energy into protons. When the RF pulse is turned off, the protons emit this extra energy and return to their original aligned positions. We refer to this procedure as relaxation. The emitted energy is detected in turn by RF coils in the scanner as the basic MR signal. To construct MR images from MR signals, we employ a system of gradient coils that can vary the strength of the magnetic field, so that each location in the brain has its own resonance frequency. The MR image is a map of these frequencies. In fact, for each voxel, the intensity of the image is proportional to the number of protons, weighted by their relaxation times in that voxel. In practice, the raw data from MRI scanner are obtained in the frequency domain. We take the inverse Fourier transform to convert it into image space, where statistical analyses are usually conducted.

Functional MRI extends the use of MRI to measure the neuronal activity of the brain indirectly and provides information on brain function. It commonly uses blood oxygenation level dependent (BOLD) contrast (Ogawa et al. 1990, Ogawa and Lee 1992). Oxygenated blood and deoxygenated blood have different magnetic susceptibility (Springer et al. 1999). Deoxygenated blood is more paramagnetic than oxygenated blood. This implies that deoxygenated blood suppress the MR signal, while oxygenated blood does not. Figure 1.1 illustrates the science of BOLD fMRI.

When the brain is active in response to a particular task or stimulus, the neurons responsible for that task increase their energy consumption, in turn, increasing oxygen demand. The change in blood oxygenation further causes changes in the local magnetic fields, which in turn affects the MR signal in the active brain regions (Kwong et al. 1992). Functional MRI has been shown to be sensitive to these changes, although it does not measure brain activity directly (Ogawa and Lee 1992).

During an fMRI experiment, the subject is placed into a MR scanner and is asked to perform a set of tasks that are designed to measure response to those tasks. The system tracks changes in fMRI signal at each voxel of the brain over multiple time points. Let n denote the number of voxels in the brain and let T represent the number of time points. Typically, n is greater than 100,000 and T varies between 100 to 2,000. Therefore, for a single subject fMRI study, we collect four-dimensional data with T multivariate observations, each with n elements. From a different perspective, the data can be considered as n time series, each with T observations. Also, other functional neuroimaging modalities, such as PET, EEG and MEG, produce the same four-dimensional data structure, although they have various spatial and temporal characteristics. In the next section, we provide a survey of statistical modeling and analysis of functional neuroimaging data. See Friston et al. (2007), Lazar (2008) and

Lindquist (2008) for more details.

1.2 Statistical Parametric Mapping

Since functional neuroimaging data tends to be very noisy due to the influences of data acquisition and physiological artifacts, a series of preprocessing steps are used to remove artifacts and clean up as much of the extraneous sources of variation as possible, prior to formal statistical analysis (Strother 2006). The major steps involved in fMRI preprocessing are slice timing correction, motion correction, registration, normalization and spatial smoothing. These steps are also crucial to validate model assumptions and make statistical analysis valid (Lindquist 2008).

Many statistical methods have been proposed for a variety of goals in functional neuroimaging data analysis (Lazar 2008). In this section, we primarily focus on the standard statistical approach to localizing brain activity, i.e. the general linear model (GLM) (Mardia et al. 1979). The GLM is at the foundation of most functional neuroimaging statistical analyses(Friston et al. 1994). The basic GLM approach assumes that voxels are independent and the same model is appropriate at each voxel in the brain. This type of analysis is called the massive univariate approach and is the most common approach. Although, recently some statistical models have been suggested that account for spatial dependence between neighboring voxels (Gossl et al. 2001, Katanoda et al. 2002, Luo and Nichols 2003, Penny et al. 2005, Bowman 2005). More specifically, for a single subject fMRI analysis, at each voxel, the fMRI time series data are fitted via a general linear model with a design matrix that reflects the stimuli present at different time points. The GLM can be extended to incorporate information for group analysis as well. Mixed-effects models have also been suggested for group analysis. After fitting the model, one performs hypothesis testing on differences in fMRI signal magnitudes between conditions or groups. This procedure is repeated for all the voxels and the test statistics (e.g. t-values) are summarized in a statistical parametric map (SPM), i.e. t-map. A thresholded SPM is usually constructed to identify regions of brain activation. In a standard fMRI study over 100,000 hypothesis tests (one for each voxel) are simultaneously made. Thus, it is important to correct for multiple comparisons to control the false positive rate. Several statistical methods have been developed based on random field theory for controlling the family-wise error rate or the false discovery rate (Genovese et al. 2002, Nichols and Hayasaka 2003). An identified brain activation region is a contiguous collection of voxels, called a cluster, whose t-values exceed a threshold for statistical significance at a pre-specified level. Most functional neuroimaging studies publish their results and only report the peak activation locations: i.e. cluster local maxima. Steps involved in a standard analysis are shown in Figure 1.3.

1.3 Meta Analysis of Functional Neuroimaging Data

As the number of functional neuroimaging studies grows, neuroscientists are increasingly interested in formal synthesis of findings across studies via meta analysis. A meta analysis mitigates the problems of a single functional neuroimaging study. First, most neuroimaging studies have relatively low power due to small sample size. The typical number of subjects in an fMRI experiment ranges from 10 to 20 due to difficulty in recruiting special patients and the high cost of fMRI scanner time. These small sample sizes are likely to provide little power to detect any signal except extremely large effects (Wager et al. 2007a, Yarkoni 2009). Meta analysis offers a way to increase statistical power by combining results from different small studies. Second, many fMRI studies published use p-values that result in high false positive rates (Wager et al. 2007b). The goal of a meta analysis is to find consistent activation region across studies making it possible to separate consistent findings from idiosyncratic ones. Third, many single functional neuroimaging studies suffer from poor reproducibility (Thirion et al. 2007). Due to the high cost of fMRI studies, replicating old studies is rarely done. Instead, new studies only retain some features of the old study. However, there is no consensus on exactly what constitute a replication. A meta analysis has the potential to address this problem by identifying consensus regions that can be tested in replications.

As pointed out in Section 1.2, single studies rarely report the full SPM. Instead, they only report foci, defined as the local maxima in significant regions in the SPM. Thus, neuroimaging meta analyses are typically based on these three-dimensional coordinates. More specifically, a typical meta analysis dataset, called coordinate-based, contains a list of foci reported from similar studies, where the number and locations of the reported foci vary from study to study. Figure 1.4 shows a meta analysis data set in both tabular and image form. Several statistical approaches have been developed to analyze coordinate-based meta analysis data (Fox et al. 1997, Nielsen and Hansen 2002, Turkeltaub et al. 2002, Wager et al. 2004, Kober et al. 2008, Eickhoff et al. 2009, Radua and Mataix-Cols 2009). The kernel-based approaches, like Activation Likelihood Estimation (ALE) and Multi-level Kernel Density Analysis (MKDA) play an important role, although they have some limitations, see Section 2.1 for a review of these methods. Such meta analysis approaches have attempted to address two main problems of interest in the field.

One is to establish activation consistency across studies to control the false positive rate and increase power. This is the most common goal of neuroimaging meta analysis. Most current methods are based on the mass-univariate approach that lacks an explicit spatial model and makes no spatial inference. This motivates the need of a fully spatial model. Thus, in this dissertation, a Bayesian spatial point process model is developed to address these issues. In many cases, the meta analysis dataset involves various psychological tasks. Neuroscientists are not only interested in identifying consistent activation regions across studies for a single task, they are also interested in the regions that are shared between the tasks. Current meta analysis methods cannot address this problem. We show that this problem can be addressed by extending the Poisson/gamma random field model (Wolpert and Ickstadt 1998a).

Another interesting problem is discriminating among the different cognitive states in a meta analysis. In fact, decoding cognitive states from human brain activity is a primary goal and a key challenge in functional neuroimaging research (Poldrack 2006, Poldrack et al. 2009). Most studies are designed to determine the distributed brain activation patterns that result from different known tasks. In a meta analysis, it is of interest to perform reverse inference on cognitive states from foci that report from a new study (Wager et al. 2009). From a statistical perspective, this can be considered a classification/prediction problem (Pereira et al. 2009). Yarkoni et al. (2011) proposed a new method. However, it can suffer from poor classification performance in some cases, see discussion in Section 4.1. One reason may be that their method only focuses on the spatial pattern of the foci and ignores the fact that the number of foci is also random. Spatial point process models do not suffer from this oversight.

1.4 Spatial Point Processes

In the section, we briefly review basic concepts and spatial point process models. A spatial point process refers to a random countable subset of a space, such as, a sub-space, $\mathcal{B}$, of d-dimensional Euclidean space, denoted by $\mathbb{R}^d$. The most important spatial point process is the Poisson point process. $\mathbf{X}$ is a Poisson point process on $\mathcal{B}$ with intensity λ, denoted by $\mathbf{X} \sim \mathcal{PP}(\mathcal{B}, \lambda)$, if and only if 1) $\forall S \subset \mathcal{B}$, $N_{\mathbf{X}}(S)$ (the number of points of $\mathbf{X}$ in S) follows a Poisson distribution with mean $\mu(S) = \int_S \lambda(x)dx$; 2) $\forall n \in \mathbb{N} = \{0, 1, 2, \ldots, \}$ and $S \subset \mathcal{B}$, conditional on $N_{\mathbf{X}}(S) = n$, there are n points of $\mathbf{X}$ in S with density $\lambda(\cdot)/\mu(S)$.

Another useful point process is the Cox process, introduced by Cox (1955) as the doubly stochastic Poisson process. For our purposes, suppose $Z(s)$ is a non-negative random field defined on the brain, $\mathcal{B} \subset \mathbb{R}^3$. If, conditional on $\lambda(s)$, $\mathbf{X}$ is a Poisson process with intensity $\lambda(s)$, then (marginally) $\mathbf{X}$ is said to be a Cox process driven by λ. If $\mathbf{X}$ is a Cox process driven by λ and λ is restricted to be a constant function, we call $\mathbf{X}$ a mixed Poisson process. Now, suppose $[\mathbf{X} \mid \lambda] \sim \mathcal{PP}(\mathcal{B}, \lambda)$. The density of this measure with respect to Lebesgue measure on $\mathcal{B}$ does not exist. However, the density (or Radon-Nikodym derivative) does exist with respect to the probability measure induced by the unit rate Poisson process defined on $\mathcal{B}$ (Møller and Waagepetersen 2004) and is

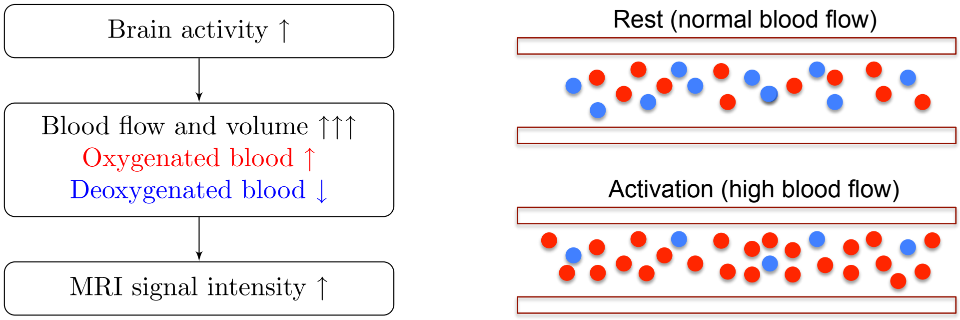
$$\pi(\mathbf{x} \mid \lambda) = \exp\left[|\mathcal{B}| - \int_{\mathcal{B}} \lambda(s)ds\right] \prod_{x \in \mathbf{x}} \lambda(x)$$

where $|\mathcal{B}|$ denotes the volume of $\mathcal{B}$. If, further, we attach a mark, $m_x \in \mathcal{M}$ to each point $x \in \mathbf{X}$, then $(\mathbf{X}, \boldsymbol{M}) = \{(x, m_x) : x \in \mathbf{X}, m_x \in \mathcal{M}\}$ is a marked Cox process with intensity $\rho : \mathcal{B} \times \mathcal{M} \to \mathbb{R}^+ \cup \{0\}$. If the M_x are independent and identically distributed with density $\pi(m)$ and are also independent of the process $\mathbf{X}$, then $\rho(s, m) = \lambda(s)\pi(m)$. Now consider a point process $\mathbf{X}$. Conditional on $\mathbf{X} = \mathbf{x}$, suppose that associated with each $x \in \mathbf{x}$ is a process $\mathbf{Y}_x$ centered at x and that these processes are independent of one another. Then $\cup_{x \in \mathbf{x}} \mathbf{Y}_x$ is an independent cluster process. Scatter noise and outliers are often modeled by a homogeneous Poisson process with intensity ϵ independently of all $\mathbf{y}_x$ (van Lieshout and Baddeley

2001). Then the independent cluster process has modified intensity $\lambda(\cdot;\mathbf{x}) = \epsilon + \sum_{x\in\mathbf{x}} f_x(\cdot;x)$. See Møller and Waagepetersen (2004) for more details on spatial point processes.

1.5 Dissertation Outline

This dissertation is organized as follows. In Chapter II, we propose a Bayesian spatial hierarchical model for functional neuroimaging meta analysis using a marked independent cluster process. In Chapter III, we develop a non-parametric Bayesian modeling approach to the analysis of multi type spatial point patterns and discuss its application to a multi type emotion meta analysis. Incorporating the models in Chapters II and III, a Bayesian spatial point process based classifier is proposed in Chapter IV for reverse inference on cognitive states in functional neuroimaging meta analysis. In the last chapter, we summarize our findings with discussions for future work.



(A) (B)

Figure 1.1: An illustration of the BOLD fMRI mechanism. (A): BOLD signal mechanism in MRI; (B): oxygenated blood (red) and deoxygenated blood (blue) flow during rest and activation.

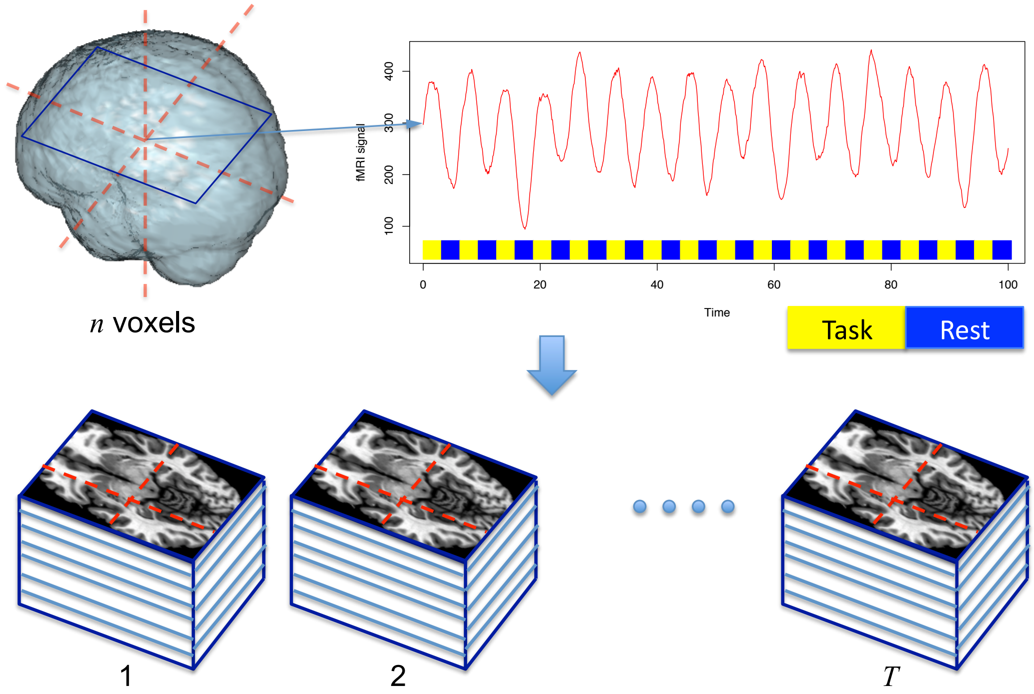
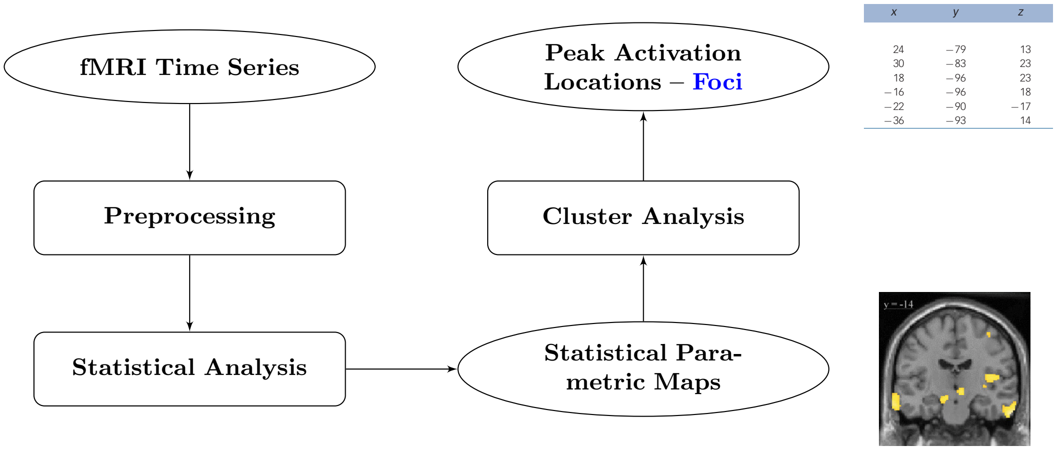


Figure 1.2: An illustration of fMRI signal and single subject data.



x	y	z
24	−79	13
30	−83	23
18	−96	23
−16	−96	18
−22	−90	−17
−36	−93	14

Figure 1.3: The fMRI data processing pipeline illustrates the paths from the raw fMRI time series to the statistical parametric maps to the peak activation locations (foci) via preprocessing, statistical analysis and cluster analysis.

Pub.	Scan	Study	Emotion	x	y	z
1	PET	1	sad	-59	-14	-1
				…	…	…
				12	-74	-35
		2	happy	48	-78	-1
				…	…	…
				12	-72	-33
		3	happy	51	-66	5
				-44	-75	15
		4	sad	-57	-16	-2
				37	26	-26
				-32	12	-17
2	PET	5	aff	48	-74	-3
				…	…	…
				-12	-69	-12
		6	aff	44	-80	-9
				…	…	…
				6	9	57
….	…	…	…	…	…	…
164	fMRI	437	mixed	-24	0	-12
				…	…	…
				48	-88	-8

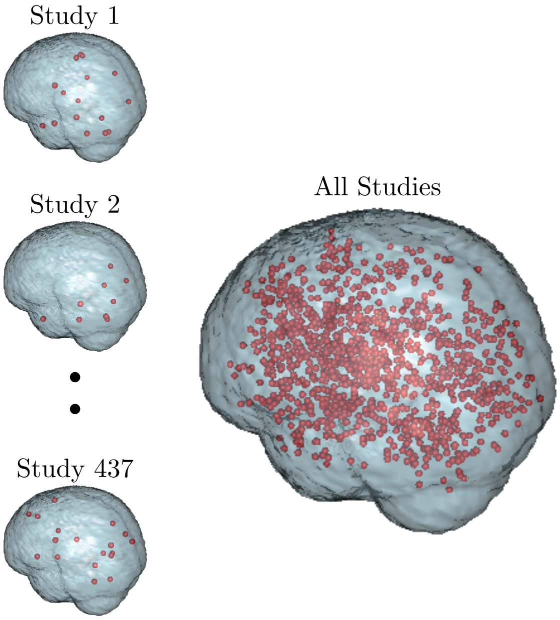


Figure 1.4: Functional neuroimaging meta analysis data: foci reported from different studies.

CHAPTER II

Meta Analysis of Functional Neuroimaging Data via Bayesian Spatial Point Processes

As the discipline of functional neuroimaging grows there is an increasing interest in meta analysis of brain imaging studies. A typical neuroimaging meta analysis collects peak activation coordinates (foci) from several studies and identifies areas of consistent activation. Most imaging meta analysis methods only produce null hypothesis inferences and do not provide an interpretable fitted model. To overcome these limitations, we propose a Bayesian spatial hierarchical model using a marked independent cluster process. We model the foci as offspring of a latent study center process, and the study centers are in turn offspring of a latent population center process. The posterior intensity function of the population center process provides inference on the location of population centers, as well as the inter-study variability of foci about the population centers. We illustrate our model with a meta analysis consisting of 437 studies from 164 publications, show how two subpopulations of studies can be compared and assess our model via sensitivity analyses and simulation studies.

2.1 Introduction

Functional neuroimaging is a relatively young discipline within the neurosciences that has led to significant advances in our understanding of the human brain (Raichle 2003). One of the most widely used method, functional magnetic resonance imaging (fMRI), has grown from just 2 publications in 1993 to over 2100 in 2009 (based on a PubMed search for “fMRI” in the title or abstract). However, due to the relatively high cost of MRI scanner time a typical fMRI study consists of fewer than 20 subjects. Thus most studies suffer from inflated type II errors (i.e. low power) and poor reproducibility (Thirion et al. 2007). See, e.g., Jezzard et al. (2001) for an overview of fMRI analysis methods.

To overcome these limitations there has been a growing interest in meta analyses of functional neuroimaging studies. The goal of a functional neuroimaging meta analysis is similar to that of a single study: to identify regions of the brain that are activated by some thought, emotion or action. Given the statistic images from the studies, or the original data, an intensity based meta analysis (IBMA) can be conducted via either a fixed effects model or an hierarchical mixed effects model (Salimi-Khorshidi et al. 2009). However, published studies rarely provide the statistic images or original data (though we note that there is a growing interest among researchers in sharing full image data and statistic maps). Rather, they only provide the locations of local maxima in the statistic image in significant regions of activation; that is, (x, y, z) coordinates in a template space, typically the Montreal Neurological Institute (MNI) template (Mazziotta et al. 2001). We shall refer to these locations as *foci*, or a single *focus*. Thus, the data in most functional neuroimaging meta analyses consist of only the foci, allowing only a coordinate based meta analysis

(CBMA). While a range of CBMA methods have been proposed, Fox et al. (1997), Nielsen and Hansen (2002), Turkeltaub et al. (2002), Wager et al. (2004), Kober et al. (2008), Eickhoff et al. (2009), Radua and Mataix-Cols (2009), we consider only the current versions of two commonly used methods, modified activation likelihood estimation (Eickhoff et al. 2009, modALE), and multilevel kernel density analysis (Kober et al. 2008, MKDA). See Salimi-Khorshidi et al. (2009) for a recent review of CBMA methods.

Both modALE and MKDA create a meta analysis statistic map based on the foci of each study, where the statistic value at each voxel (or volume element) summarizes the evidence for clustering at that location. Briefly, they start by creating an image for each focus, where the intensity in the image is based on the proximity of each voxel to the focus. These per-focus maps are then combined into a study map, and the study maps are in turn combined into a meta analysis map. The intensities in the meta analysis map are compared to maps generated by null hypothesis Monte Carlo simulation, creating P-values. The two methods differ in how they create the foci maps, and how they combine these into study and meta analysis maps. The modALE creates a focus map by placing a Gaussian density of size $\sigma_{\mathrm{mod}ALE}$ centered at the focus (normalized to integrate to unity over the map), where the intensity is interpreted as the probability that the focus arose from a given voxel. Assuming independence between foci, modALE combines focus maps with the probability addition rule, effectively computing the probability that one or more foci arose at a given voxel; this procedure is used both to combine focus maps into study maps, and to combine study maps into meta analysis maps. MKDA creates a focus map by placing a sphere of unit intensity and diameter $d_{\mathrm{M}DKA}$ centered at the focus. Multiple focus maps are combined into a study map with the logical OR operator, creating an indicator map showing where one or more foci are found within distance $d_{\mathrm{M}DKA}$ of a given voxel. Multiple study maps are combined into a meta analysis map with the sample mean, providing a map interpretable as the proportion of studies with one or more foci within distance $d_{\mathrm{M}DKA}$. Monte Carlo resampling is roughly similar in each method, with foci locations randomly shuffled between studies, and meta analysis maps recreated for each realization; see the respective references for details. Traditional mass-univariate statistical inference is carried out, finding either voxel-wise or cluster-wise P-values, corrected by familywise-error methods or the false discovery rate.

While methods like modALE and MKDA are widely used, have freely available implementations, and, for MKDA in particular, have intuitive appeal, we find that they have several shortcomings. Both require spatial kernel parameters ($\sigma_{\mathrm{mod}ALE}$ and $d_{\mathrm{M}DKA}$) that must be fixed; while Eickhoff et al. (2009) propose an algorithm for estimating $\sigma_{\mathrm{mod}ALE}$ based on training data, the value is not data adaptive and is assumed to be constant over the brain. Further, while neuroimaging users are familiar with the voxel-wise and cluster-wise inferences generated, these are based on a mass-univariate approach that lacks an explicit spatial model. Specifically, there is no way to infer on spatial dispersion of foci between studies, nor obtain spatial confidence regions on where foci arise in the population of all studies.

In this work we propose a hierarchical spatial point process model (Møller and Waagepetersen 2004; 2007, Illian et al. 2008) and estimate model parameters via the Bayesian paradigm. In particular, we adopt a spatial independent cluster process (van Lieshout and Baddeley 2001). The algorithm used to estimate the posterior distribution of model parameters stochastically searches for clusters of foci. Clusters appear in regions of (relatively) high foci density across studies and thus represent regions where a preponderance of studies reported activation. A cluster's center represents the most likely location of the foci that define the cluster, and thus represents the "population center" of activation in the particular region of the brain in which the cluster is observed. Thus, a central goal of our modeling is to find clusters of foci and their associated population centers. Furthermore, since we adopt a Bayesian modeling approach, other quantitative information can be extracted from our model that cannot be deduced from current CBMA methods. Such as 1) the variability of the foci about the population center; 2) the variability of the population centers themselves; 3) the probability that there exists at least one population center in any region of interest (ROI) within the brain; 4) the probabilistic comparison of locations of cluster (population) centers of foci across different types of studies (e.g. studies of negative vs. positive emotion); 5) prediction of where a new study will most likely report foci (and hence the most likely locations where activation will be found); and 6) estimation of the proportion of foci that do not cluster with foci from other studies. Specific examples are given in Section 2.3.

To give an impression of the neuroimaging meta analysis data, Figure 2.1 shows an extract of the foci data of the emotion meta analysis in both tabular and image form (see Section 2.3 for more details). Note that both PET and fMRI data are considered, reasonable since, after smoothing, group fMRI studies are similar to PET data in smoothness and interpretation of the signal (Feng et al. 2004). An important facet of the data is the issue of *singly reported foci* versus *multiply reported foci*. For a given activation area in the brain, some authors only report a single focus, while others report multiple foci, however this information is rarely provided in the literature. These differences are attributable to how different software packages report results, and simply author preference.

The remainder of this manuscript is organized as follows. In Section 2.2 we propose our Bayesian hierarchical marked spatial independent cluster process. In Section 2.3, we apply our model to a meta analysis data set of emotion studies and compare our results from modALE and MKDA. In Section 2.4 we briefly discuss results from sensitivity and simulation studies. We conclude the paper with a discussion of our model and ideas for future research. We provide supplementary material in Appendix A with algorithm details, pseudo code, and details from simulation studies and sensitivity analyses.

2.2 The Model

2.2.1 Model Outline and Notation

First we outline our model and then present details. The hierarchical model consists of three main levels and is illustrated in Figure 2.2. At the lowest level, level 1, are the foci (data). For study c, $c = 1, \dots, C$, the foci are a realization, $\mathbf{x}_c$, of an independent cluster process, $\mathbf{X}_c$, driven by random intensity function λ_{1c} and these processes are independent across studies. The process $\mathbf{X}_c$ is made up of two types of foci: singly reported foci (type 0) and multiply reported foci (type 1). For a generic point $x \in \mathbf{x}_c$ denote the missing type indicator, or mark, by δ_x. Let $\mathbf{X}_c^d = \{x \in \mathbf{X}_c : \delta_x = d\}$, $d = 0, 1$. Conditional on the realization, $\mathbf{y}_c$, of a latent *study activation center* process, $\mathbf{Y}_c$, we associate with each $y \in \mathbf{y}_c$ a process, $\mathbf{X}_{cy}^1$, of type 1 foci normally distributed about the study activation center y with covariance $\mathbf{\Psi}_y$ and that these processes are independent and their union $\mathbf{X}_c^1 = \cup_{y \in \mathbf{y}_c} \mathbf{X}_{cy}^1$ forms an independent cluster process driven by random intensity function λ_{1c}^1. Note that $\mathbf{\Psi}_y$ is a latent, random mark attached to $y \in \mathbf{y}_c$. Also, conditional on the realization, $\mathbf{z}$, of a latent *population center* process, $\mathbf{Z}$, we associate with each $z \in \mathbf{z}$ a process, $\mathbf{X}_{cy}^0$, of type 0 foci normally distributed about the population center z with covariance

$\mathbf{\Sigma}_z$ (a latent random mark attached to z) and that these processes are independent across studies. We also allow the possibility that type 0 foci do not cluster about any $z \in \mathbf{z}$ and model them as an independent homogeneous Cox process $\mathbf{X}_{c\emptyset}$ driven by homogeneous random intensity ϵ_{1c}. The union $\mathbf{X}_c^0 = \cup_{z\in\mathbf{z}}\mathbf{X}_{cz}^0 \cup \mathbf{X}_{c\emptyset}$ forms an independent cluster process driven by random intensity function λ_{1c}^0. We note that $\mathbf{X}_c = \mathbf{X}_c^0 \cup \mathbf{X}_c^1$ and is driven by the random intensity $\lambda_{1c} = \lambda_{1c}^0 + \lambda_{1c}^1$.

At level 2 we model the latent study activation center process for study c, $\mathbf{Y}_c$, as an independent cluster process. Conditional on the (realized) population center process, $\mathbf{z}$, we associate with each z a finite process, $\mathbf{Y}_{cz}$, with realization $\mathbf{y}_{cz}$ of points normally distributed about the population center z with covariance $\mathbf{\Sigma}_z$. We assume that the processes $\mathbf{Y}_{cz}$ given $\mathbf{z}$ are independent. We allow the possibility that some study activation centers do not cluster about any $z \in \mathbf{z}$ and model these as an independent homogeneous Cox process $\mathbf{Y}_{c\emptyset}$ driven by homogeneous random intensity ϵ_{2c}. and thus their union $\mathbf{Y}_c = \cup_{z\in\mathbf{z}}\mathbf{Y}_{cz} \cup \mathbf{Y}_{c\emptyset}$ forms an independent cluster process driven by the random intensity function λ_{2c}. (A study activation center is a substitute for the location of the global maximum in a given activation region, which is not reported. This is how we account for multiply reported foci.)

At the highest level, level 3, are the latent population centers. We assume, a priori, that the population centers are a realization, $\mathbf{z}$, of a homogeneous Cox process $\mathbf{Z}$ driven by homogeneous random intensity β. Attached to each population center, $z \in \mathbf{z}$, is a latent random mark, $\mathbf{\Sigma}_z$. The points that cluster about the population centers are singly reported foci from level 1 and study activation centers from level 2. We refer to the singly reported foci and the study activation centers collectively as *activation centers*.

2.2.2 Model Details

We begin at level 1. We model $\mathbf{X}_c$ with an independent cluster process on the brain, $\mathcal{B}$, driven by $\lambda_{1c}(x;\cdot)$. (Throughout, $f(x;\cdot)$ represents a parametrized function f of x. The '$\cdot$' is shorthand for all parameters on which f depends.) The conditional likelihood of this process for study c is

$$\pi\left[\mathbf{x}_c \mid \lambda_{1c}(x;\cdot)\right] \propto \exp\left[-\int_{\mathcal{B}} \lambda_{1c}(s;\cdot)ds\right] \prod_{x\in\mathbf{x}_c} \lambda_{1c}(x;\cdot). \tag{2.1}$$

We note here that $\mathbf{X}_c \mid \lambda_{1c}(x;\cdot)$ is a Poisson point process and that the conditional likelihood defined in (2.1) is the density of $\mathbf{X}_c \mid \lambda_{1c}(x;\cdot)$ with respect to the measure of a unit-rate Poisson process and not to the standard Lesbesgue measure (Møller and Waagepetersen 2004; 2007). The unmarked process $\mathbf{X}_c$ is made up of the two types of foci—singly reported, type 0, and multiply reported, type 1. For a generic point $x \in \mathbf{X}_c$ assume, a priori, that $\pi(\delta_x = 0) = p = 1 - \pi(\delta_x = 1)$. Given these marks the processes $\mathbf{X}_c^0$ and $\mathbf{X}_c^1$ are independent and are driven by intensity functions λ_{1c}^0 and λ_{1c}^1, respectively, with $\lambda_{1c} = \lambda_{1c}^0 + \lambda_{1c}^1$. Let $(\mathbf{x},\boldsymbol{\delta})_c = \{(x,\delta_x) : x \in \mathbf{x}_c\}$. The joint density of the data and marks is

$$\begin{aligned}
\pi\left[(\mathbf{x},\boldsymbol{\delta})_c \mid \lambda_{1c}\right] &= \prod_{d=0}^{1} \pi\left[\mathbf{x}_c^d \mid \lambda_{1c}\right] \prod_{x\in\mathbf{x}_c} p^{1-\delta_x}(1-p)^{\delta_x} \\
&\propto \exp\left[-\int_{\mathcal{B}} \lambda_{1c}^0(s;\cdot)ds\right] \prod_{x\in\mathbf{x}_c} \left[\lambda_{1c}^0(x;\cdot)p\right]^{1-\delta_x} \times \\
&\qquad \exp\left[-\int_{\mathcal{B}} \lambda_{1c}^1(s;\cdot)ds\right] \prod_{x\in\mathbf{x}_c} \left[\lambda_{1c}^1(x;\cdot)(1-p)\right]^{\delta_x}.
\end{aligned} \tag{2.2}$$

Now (2.2) and (2.1) are equivalent in the sense that if we marginalize over all possible $\boldsymbol{\delta}_c = \{\delta_x : x \in \mathbf{x}_c\}$ in (2.2) we get (2.1). The intensity functions are

$$\lambda_{1c}^0(x;\cdot) = \epsilon_{1c} + \sum_{z\in\mathbf{z}} \theta_{1c}\phi_3(x;z,\boldsymbol{\Sigma}_z) \tag{2.3}$$

$$\lambda_{1c}^1(x;\cdot) = \sum_{y\in\mathbf{y}_c} \eta_c\phi_3(x;y,\boldsymbol{\Psi}_y). \tag{2.4}$$

The function $\phi_3(a; b, \boldsymbol{A}_b)$ represents the density of a 3-D normal random variable with mean b, covariance matrix $\boldsymbol{A}_b$ at location a. The parameter θ_{1c} multiplied by $\int_{\mathcal{B}} \phi_3(x; z, \boldsymbol{\Sigma}_z)dx$ is the expected number of type 0 foci that cluster about population center $z \in \mathbf{z}$, while $\eta_c \int_{\mathcal{B}} \phi_3(x; y, \boldsymbol{\Psi}_y)dx$ is the expected number of type 1 foci that cluster about study activation center $y \in \mathbf{y}_c$. The intensity function of the type 0 foci that do not cluster about a population center is ϵ_{1c}.

Level 2: Let $(\mathbf{y}, \boldsymbol{\Psi})_c = \{(y, \boldsymbol{\Psi}_y), y \in \mathbf{y}_c\}$. The joint density of the conditional latent study activation center process and the independent marking distribution, for study c, is

$$\begin{aligned}\pi\left[(\mathbf{y}, \boldsymbol{\Psi})_c \mid \lambda_{2c}\right] &= \pi\left[\mathbf{y}_c \mid \lambda_{2c}(y; \cdot)\right] \prod_{y \in \mathbf{y}_c} \pi\left[\boldsymbol{\Psi}_y\right] \\ &\propto \exp\left[-\int_{\mathcal{B}} \lambda_{2c}(s; \cdot)ds\right] \prod_{y \in \mathbf{y}_c} \lambda_{2c}(y; \cdot)\pi(\boldsymbol{\Psi}_y).\end{aligned}$$

The intensity function of the unmarked process $\mathbf{Y}_c$ is given by

$$\lambda_{2c}(y; \cdot) = \epsilon_{2c} + \sum_{z \in \mathbf{z}} \theta_{2c}\phi_3(y; z, \boldsymbol{\Sigma}_z). \tag{2.5}$$

In (2.5), $\theta_{2c} \int_{\mathcal{B}} \phi_3(y; z, \boldsymbol{\Sigma}_z)dy$ is the expected number of study activation centers that cluster about population center $z \in \mathbf{z}$. The intensity function of the study activation centers that do not cluster about a population center is ϵ_{2c}.

By independence the joint density of the processes $\mathbf{X}_c^0$, $\mathbf{Y}_c$, $c = 1, \ldots, C$, is

$$\prod_{c=1}^{C} \pi(\mathbf{x}_c^0 \mid \lambda_{1c}^0)\pi(\mathbf{y}_c \mid \lambda_{2c}) \propto \exp\left(-\int_{\mathcal{B}} \lambda(s; \cdot)ds\right) \prod_{c=1}^{C} \prod_{x \in \mathbf{x}_c^0} \lambda_{1c}^0(x; \cdot) \prod_{y \in \mathbf{y}_c} \lambda_{2c}(y; \cdot)$$

where

$$\begin{aligned}\lambda(y; \cdot) &= \left(\sum_{c=1}^{C} (\epsilon_{1c} + \epsilon_{2c})\right) + \left(\sum_{c=1}^{C} (\theta_{1c} + \theta_{2c})\right) \sum_{z \in \mathbf{z}} \phi_3(y; z, \boldsymbol{\Sigma}_z) \\ &\equiv \epsilon + \theta \sum_{z \in \mathbf{z}} \phi_3(y; z, \boldsymbol{\Sigma}_z). \qquad (2.6)\end{aligned}$$

Level 3: At the final level, the latent, unmarked, population center process $\mathbf{Z}$ is modeled as a homogeneous Cox process driven by β defined on $\mathcal{B}$. Let $|\mathcal{B}|$ denote the volume of $\mathcal{B}$. Let $(\mathbf{z}, \boldsymbol{\Sigma}) = \{(z, \boldsymbol{\Sigma}_z), z \in \mathbf{z}\}$. The conditional joint density of the population center process and the independent marking distribution is

$$\pi\left[(\mathbf{z}, \boldsymbol{\Sigma}) \mid \beta, \boldsymbol{T}\right] = \pi(\mathbf{z} \mid \beta) \prod_{z \in \mathbf{z}} \pi(\boldsymbol{\Sigma}_z \mid \boldsymbol{T}) \propto \exp\left(-\beta|\mathcal{B}|\right) \prod_{z \in \mathbf{z}} \beta\, \pi(\boldsymbol{\Sigma}_z \mid \boldsymbol{T})$$

where $\boldsymbol{T}$ is a hyperprior for the distribution of $\boldsymbol{\Sigma}_z$.

We now specify prior distributions and begin with level 3 priors and work backwards. For $z \in \mathbf{Z}$, $\boldsymbol{\Sigma}_z \sim W^{-1}(\boldsymbol{T}, \nu)$; that is, $\boldsymbol{\Sigma}_z$ has an inverse Wishart distribution with scale matrix $\boldsymbol{T}$ and ν degrees of freedom. The $\boldsymbol{\Sigma}_z$ are independent of one another and are independent of the process $\mathbf{Z}$. The inverse of the hyperparameter $\boldsymbol{T}$ is assigned a Wishart distribution: $\boldsymbol{T}^{-1} \sim W(\boldsymbol{T}_0, \nu_0)$ where $\boldsymbol{T}_0$ and ν_0 are fixed. The random intensity, β, of $\mathbf{Z}$ is assigned, a priori, a gamma distribution: $\beta \sim G(a_\beta, b_\beta)$ with a_β and b_β fixed (values of the fixed hyper-parameters should be problem specific and are discussed below in Section 2.3). Now at level 2, the marks, $\boldsymbol{\Psi}_y$, of the study activation center processes are given an inverse Wishart distribution: $\boldsymbol{\Psi}_y \sim W^{-1}(\boldsymbol{S}, d)$. The $\boldsymbol{\Psi}_y$ are independent of one another and independent of the processes $\mathbf{Y}_c$, $c = 1, \ldots, C$. Both ϵ and θ defined in (2.6) are assumed known. In the simulation of the posterior distribution, it is not necessary to estimate the parameters ϵ_{2c} and θ_{2c}, $c = 1, \ldots, C$ and estimates of ϵ_{1c} and θ_{1c} are only needed to impute the missing type indicator (whether a focus is a singly reported focus or a multiply reported focus). We assume the probability that a type 0 focus in study c clusters about a population center $z \in \mathbf{z}$ and the probability that a study activation center in study c clusters about the same population center are equal. We feel this assumption is quite reasonable as it implies that the study activation centers and the singly reported foci are treated equivalently in level 2 (see the intensity functions in (2.3), (2.5) and (2.6)). Furthermore, we assume that these probabilities are also equal across studies. This assumption also reduces the number of parameters that need to be estimated. In Appendix A, we show that this *probability equivalence assumption* implies $\theta_{1c}/\epsilon_{1c} = \theta_{2c}/\epsilon_{2c} = \theta/\epsilon$, which, in turn, implies that $\epsilon_{1c}/\epsilon = \theta_{1c}/\theta \, (\equiv \rho_{1c})$ and $\epsilon_{2c}/\epsilon = \theta_{2c}/\theta \, (\equiv \rho_{2c})$. Thus, $\sum_{c=1}^{C}(\rho_{1c} + \rho_{2c}) = 1$. Define $\boldsymbol{\rho} = (\rho_{11}, \ldots, \rho_{1C}, \rho_{21}, \ldots, \rho_{2C})$. A priori, we assign $\boldsymbol{\rho}$ a Dirichlet distribution: $\boldsymbol{\rho} \sim D(\alpha_1, \ldots, \alpha_1, \alpha_2, \ldots, \alpha_2)$. The prior distribution on $\boldsymbol{\rho}$ induces a prior distribution on the ϵ_{1c}, θ_{1c}, ϵ_{2c} and the θ_{2c}. The last level 1 parameters are the η_c. A priori, we assume that $\eta_c \sim G(a_\eta, b_\eta)$ and are independent of one another.

The posterior distribution of parameters given data is complicated and has no closed form solution. Thus we resort to spatial birth and death processes nested within a Markov chain Monte Carlo simulation algorithm to sample from the posterior distribution. Details of the algorithm and pseudo code are provided in Appendix A.

2.3 Application

In this section, we apply our model to a neuroimaging meta analysis of emotion first reported in Kober et al. (2008). The meta analysis data set consists of 164 publications of various aspects of emotion. A total of 7 emotions were studied across the different experiments: sad, happy, anger, fear, disgust, surprise and affective. Our goal is to find consistent regions of activation across the different studies and types of emotions. Many papers report results from different statistical comparisons called “contrasts”, which we are calling studies. Following the convention of existing neuroimaging meta analyses, we treat each of these studies as independent. There are a total of 437 studies reporting a total of 2475 foci. Table 2.1 lists some features and summary statistics of this data set. Consult Kober et al. (2008) for further details.

2.3.1 Prior Parameters

Several parameters are assigned vague or non-informative prior distributions. We assign a vague prior to the η_c: $\eta_c \sim G(0.001, 0.001)$, $c = 1, \ldots, C$, while $\boldsymbol{\rho}$ is assigned the non-informative Jeffrey's prior: $\boldsymbol{\rho} \sim D(0.5, \ldots, 0.5, 0.5, \ldots, 0.5)$. All other prior and hyperprior parameters are obtained by elicitation from an expert in the meta analysis of neuroimaging data, and neuropsychologist, Tor Wager. We asked Tor, based on his experience, 5 questions: i) How many population centers do you expect to find in this meta analysis? ii) Given that some studies report multiply reported peaks per activation regions and other do not, on average how many multiply reported peaks per study do you expect? or how many activation centers per study do you expect? iii) What percentage of the activation centers do you expect to cluster about population centers? iv) What is the average spread of multiply reported foci about study activation centers? v) What is the average spread of activation centers about population centers? We note that since we need to match expected numbers, or percentages, given in these responses to the actual data, some prior settings have an empirical Bayesian flavor. Given his responses, we derive the remaining prior distributions as follows.

i) The number of population centers, or clusters, for this particular meta analysis is in the range from 20 to 40. Thus, a priori, we set the expected number of population centers, $E\left[N_{\mathbf{Z}}(\mathcal{B})\right]$, to 30. We want to be vague about the range of the number of population centers and thus set $\beta|\mathcal{B}| \sim G(0.03, 0.001)$. Therefore, since $[N_{\mathbf{Z}}(\mathcal{B}) \mid \beta|\mathcal{B}|]$ has a Poisson distribution with mean $\beta|\mathcal{B}|$, $\underline{N_{\mathbf{Z}}(\mathcal{B})}$ is, a priori, a negative binomial random variable with mean 30 and variance $30,030$. ii) The mean number of foci reported per study is 5.67 (Table 2.1). We expect that there will be on average, 5 singly reported foci and study activation centers (collectively, activation centers) per study (for a total of 2185) and that iii) the majority of these will cluster about population centers—80% (for a total of 1748). Let $\mathbf{A} = \cup_{c=1}^{C} \cup_{z \in \mathbf{z}} (\mathbf{X}_{cz}^{0} \cup \mathbf{Y}_{cz})$, then $E(N_{\mathbf{A}}(\mathcal{B})) \equiv \theta \sum_{z \in \mathbf{z}} \Phi_3(\cdot; z, \mathbf{\Sigma}_z) = \theta \sum_{z \in \mathbf{z}} \int_{\mathcal{B}} \phi_3(\xi; z, \mathbf{\Sigma}_z) d\xi$ is the expected number of activation centers that cluster about population centers; i.e., conditional on θ, $\mathbf{z}$ and the marks $\mathbf{\Sigma}_z, z \in \mathbf{z}$, $N_{\boldsymbol{A}}(\mathcal{B})$ is a Poisson random variable with mean $\theta \sum_{z \in \mathbf{z}} \Phi_3(\cdot; z, \mathbf{\Sigma}_z)$. Equating the latent number of population centers to the mean number of population centers, $n_{\mathbf{z}}(\mathcal{B}) = E(N_{\mathbf{Z}}(\mathcal{B})) = 30$, and assuming $\Phi_3(\cdot; z, \mathbf{\Sigma}_z) \approx 1, \forall z \in \mathbf{z}$, we have $E(N_{\mathbf{A}}(\mathcal{B})) \approx 30\theta = 1748$ which implies that $\underline{\theta = 1748/30}$. Also, $\epsilon|\mathcal{B}|$ is the expected number of activation centers that do not cluster about population centers; i.e. $\mathbf{A}_{\emptyset} \equiv \cup_{c=1}^{C} (\mathbf{X}_{c\emptyset} \cup \mathbf{Y}_{c\emptyset}) \sim \text{Poisson}(\mathcal{B}, \epsilon)$ so that $N_{\mathbf{A}_{\emptyset}}(\mathcal{B})$ is a Poisson random variable with mean $\epsilon|\mathcal{B}| = 437$. Thus, $\underline{\epsilon = 437/|B|}$. iv) The covariances, or marks, $\underline{\mathbf{\Psi}_y \sim W^{-1}(\boldsymbol{S}, 5)}$ where S is the 3×3 identity matrix, $\boldsymbol{I}$. This gives, a priori, $E(\mathbf{\Psi}_y) = \boldsymbol{I}$. v) The covariances $\underline{\mathbf{\Sigma}_z \sim W^{-1}(\boldsymbol{T}, 5)}$ with $\underline{\boldsymbol{T}^{-1} \sim W(\boldsymbol{T}_0, 5)}$ where $\boldsymbol{T}_0 = 0.8\boldsymbol{I}$ which results in, a priori, $E(\mathbf{\Sigma}_z) = 4\boldsymbol{I}$. We note here that if $\boldsymbol{A} \sim W^{-1}(\boldsymbol{B}, d)$ and $\boldsymbol{A}$ has dimension $m \times m$, that the variance and covariances of the elements of $\boldsymbol{A}$ do not exist when $d \leq m + 3$ (Press 1982). Thus, the prior distributions of $\mathbf{\Psi}_y$ and $\mathbf{\Sigma}_z$ are heavy-tailed.

A sensitivity analysis of the posterior distribution to the informative prior information in our model (numbers 1 through 5 above; i.e. $N_{\mathbf{Z}}(\mathcal{B})$, ϵ, θ, $\mathbf{\Psi}_y$ and $\mathbf{\Sigma}_z$ and hyperprior $\boldsymbol{T}$) is provided in Appendix A. We briefly discuss our findings in Section 2.4, below. Next, we present results from our modeling of the emotion meta analysis dataset.

2.3.2 Analysis of the Emotion Meta Analysis Dataset

We approximate the posterior distribution by running the algorithm for 120,000 iterations, discarding the first $20,000$ as a burn-in. We assess convergence of the chain by multiple runs of the algorithm from diverse initial conditions and visually inspect the difference in various posterior mean intensity functions and find only minor differences. Furthermore, we use the method of Gelman and Rubin (1992) to assess convergence on the number of population centers. The mean of the potential scale reduction factor is 1.0 with an upper 0.975 quantile of 1.01. Thus, the number of iterations and burn-in appears to be sufficient and that the chain has converged to stationarity.

First we compare our results with those from a modALE (Eickhoff et al. 2009) and a MKDA (Wager et al. 2007b, Kober et al. 2008) analysis of the same data. Since we do not have an auxiliary data set that can be used to estimate the kernel size used in modALE, we use the default kernel size provided in the software. In fact, the software does not allow the user to define the kernel size. To the best of our knowledge, the kernel size defined in the software is that derived in Eickhoff et al. (2009) and is based on a fist clenching experiment which may not be appropriate for our data. Figure 2.3 shows a visual, qualitative comparison of the activation center and population center intensity functions and the modALE and MKDA maps for 11 equally spaced, 2 mm axial slices, throughout the brain. Although qualitatively similar, there are visible differences. For instance, in the third column of Figure 2.3 the activation center intensity from our model appears to be more concentrated than either the modALE or MKDA map. We also note that we have separated out two intensity functions: the activation center intensity function and the population center intensity function.

We can identify several regions of high intensity, with the highest intensities centered in slice $Z = -22$mm. These bilateral regions are the amygdalae and the high activation center intensity indicates a preponderance of studies clustering in these regions and that the clustering is tight. Both modALE and MKDA also identify these regions with large statistic values. We find very high, very concentrated intensity in each amygdala in the population center intensity function as well. (The population center intensity function is created by smoothing the posterior histogram of population centers with a Bayesian nonparametric density estimation model. In particular, a mixture of Dirichlet process priors model (Escobar and West 1995).) This indicates that the variability of the population centers is much smaller than the variability of the activation centers about these population centers. Quantitatively, however, it is difficult to compare results between our model and CBMA methods, as they have very different interpretations. At a particular voxel, the value in the modALE map is the probability that at least one focus occurs at said voxel across studies. The value in the MKDA map has the interpretation of a (weighted) proportion of studies that report a focus within a prespecified distance to that voxel. Whereas our posterior intensity functions are interpreted as just that: intensities of activation centers or population centers given the data. Given a voxel, of say volume v, the integrated intensity over the volume of the voxel (or over any ROI) is interpreted as the expected number of activation centers (or population centers, as the case may be) in that voxel or ROI. Note that if the intensity function is normalized by its integral over the brain, then the normalized intensity function can be interpreted as a spatial density function. The integral of the spatial density function over any ROI is the probability of an activation (population) center occurring in that ROI. There is a distinct difference between this interpretation and that of modALE. In modALE, the probabilities are per voxel so that the probability measure integrates to one at each voxel. Thus, modALE is a massive univariate approach, as is MKDA. Whereas, the spatial density function integrates to one over the entire brain and not at each voxel.

Computationally, modALE takes around 20 minutes, MKDA takes around 2 hours (1000 permutations) and our model takes approximately 20 hours to sample from the posterior (120K iterations), all on a 2.4 GHz iMAC.

Since we use a Bayesian hierarchical model, extra information can be extracted than from current CBMA methods. The information that can be extracted is enumerated in the introduction and we now provide examples.

1) and 2). Given any prespecified ROI, our model provides location and variability of location of activation centers about population centers and about the variability of population centers, themselves, within the ROI. We use an amygdala ROI, as over 50 years of research has implicated its role in emotion (see Phelps and LeDoux (2005) for a review). For example, integrating over the z-dimension, the posterior 95% credible ellipses of both population centers and activation centers within the amygdalae are provided in Figure 2.4. These credible ellipses represent the uncertainty in the location of the population centers that are found in the amygdalae (the gray ellipses) and the variability of activation centers that cluster about the population centers that are found in the these structures (the larger, white ellipses). Note that these latter ellipses are conditional on the event that a single population center occurs in the respective amygdala. A single population center occurred in 69% of the iterations in the left amygdala and 90% in the right.

3) At lease one population center occurred in over 99.9% of the iterations for both amygdalae. In Appendix A, Table 4, we provide the probability that at least one population center occurs in various ROIs.

The volumes of the ellipsoids are given in Table 2.2 and quantify the inter-study variability of activation centers about the population centers and the variability in the locations of the population centers. The location of the maximum statistic value from modALE and MKDA, conditional on being located in each amygdala, are also provided in Table 2.2 for comparison. Although each activation center ellipsoid is about 3 times the volume of the respective amygdala, any activation center within one of these ellipsoids, is deemed close enough to be associated with the amygdala, irrespective of our model. That is, a neuropsychologist would consider them to be "associated with, or part of, the amygdalae". The volume of the 95% credible ellipsoids of population centers are roughly 100 times smaller than the corresponding activation center ellipsoids. This demonstrates a key strength of our Bayesian hierarchical model: the ability to quantify the precision of the population locations and to quantify the precision of the activation centers as they cluster about population centers. We note here that there is an identifiability issue with the intensity function defined in (2.6): it is invariant to permutations of the indices in the summations, and thus the intensity is invariant as well. However, by conditioning on the event that exactly one population center occurs in, say, the right amygdala, we conditionally removed the lack of identifiability.

To the best of our knowledge, our model is the first model of neuroimaging meta analysis data that quantifies these precisions and that separates the precision of the activation centers about population centers and the precision of the location of population centers. CBMA methods, such as ALE and MKDA, provide point estimation of activation regions and do not quantify the associated estimation error. (A bootstrap estimate of standard errors, of the modALE or MKDA map, is conceivable, however the computational cost would be large—for a bootstrap sample of size n, roughly n times longer than a single run—and would not allow separation of sources of variability.)

In Figure 2.3 we also provide the posterior predictive intensity. This function provides information about where a new, future, study of emotion would most likely report activation centers. For example, the expected number of activation centers in a new study is 5.62. Integrating the predictive intensity over each amygdala results in an expected number of activation centers of 0.090 and 0.092 in the right, left amygdala, respectively.

4) To demonstrate how our model can be used to compare subpopulations of studies, we split our meta analysis into studies based on positive emotions and those based on negative emotions. In particular, there is interest in whether the brain regions that subserve positive and negative emotions are the same. Specifically, is the location of activation the same for both types of emotional stimuli within an amygdala? To address this question we apply our model to the positive and negative emotions subsets. Convergence was assessed visually and by computing the multivariate potential scale reduction factor for both the location of the left (upper bound of reduction factor = 1.02) and right (upper bound = 1.01) amygdala population centers.

There are 522 foci from 95 studies of positive emotions and 1663 foci from 281 studies of negative emotions. For each amygdala, let Z_p and Z_n denote the positive and negative emotions population centers, respectively, located in the amygdala, conditional on the event that there is exactly one population center in the amgydala. The estimated posterior distribution of Z_p and Z_n can be approximated by normal distributions: $Z_p \sim N(\mu_p, \Sigma_p)$ and $Z_n \sim N(\mu_n, \Sigma_n)$, where μ_p and μ_n are the mean locations of the population centers and Σ_p and Σ_n are the covariance matrices, from which we can compute the associated 95% credible ellipsoids (see Figure 2.5). With the assumption that positive and negative emotions studies are independent, the difference in locations is $Z_d = Z_p - Z_n \sim N(\mu_p - \mu_n, \Sigma_p + \Sigma_n)$. Thus we can compute the 95% credible ellipsoid of Z_d for each amygdala. In both the left and right amygdala, the 95% credible ellipsoid excludes the origin—indicating a substantial difference in location. We also estimate the posterior distribution of the Euclidean distance between Z_p and Z_n, i.e. $E_{pn} = \sqrt{(Z_p - Z_n)'(Z_p - Z_n)}$, from which we can estimate the probability $\Pr(E_{pn} > d)$ for different d's. For example, in the right amygdala $\Pr(E_{pn} > 2\text{mm}) > 0.999$ and $\Pr(E_{pn} > 4\text{mm}) = 0.932$. For the left amygdala $\Pr(E_{pn} > 2\text{mm}) = 0.983$ and $\Pr(E_{pn} > 4\text{mm}) = 0.704$. This analysis suggests there is strong evidence of a difference in location between the positive and negative emotions population centers located in the right amygdala, and modest evidence for a difference in the left amygdala. To our knowledge, our model is the first one to be able to quantify and draw inferences on differences in population locations between studies.

5) The posterior predictive intensity of a new study is shown in the fourth row of Figure 2.3. It is qualitatively similar to the posterior activation center intensity, however some minor differences can be seen.

6) Lastly, the posterior mean of the proportion of activation centers that do not cluster about any population center is 0.22 with a standard deviation of 0.01.

2.3.3 Model Assessment

We conduct a posterior predictive model assessment using the L function which is a summary statistic for second order properties of a point process (Baddeley et al. 2000, Illian et al. 2009). The L function can indicate aggregation or clustering for a point process. For our model, $L(r;\cdot) = \{3K(r;\cdot)/4\pi\}^{1/3}$, where

$$K(r;\mathbf{X}_c,\cdot) = \frac{1}{|\mathcal{B}|}\sum_{x_1,x_2\in\mathbf{X}_c}\frac{\mathbf{1}[\|x_1-x_2\|\leq r]}{\lambda_{1c}(x_1;\cdot)\lambda_{1c}(x_2;\cdot)}$$

Consider the posterior predictive distribution of the differences $\Delta_c(r) = L(r;\mathbf{X}_c,\cdot) - L(r;\mathbf{X}_c^*,\cdot)$, where $\mathbf{X}_c^*$ is a simulated sample from the posterior predictive distribution. As discussed by Illian et al (2009), if zero is an extreme value in the posterior predictive distribution of $\Delta_c(r)$ for a range of distances r, then we may question the fit of our model. We estimate the upper and lower boundaries of the 95% posterior intervals for the posterior predictive distributions of $\Delta_c(r), r > 0$, for the 437 studies ($c = 1,\ldots,437$). Over ninety percent (395/437) of the studies have 95% posterior intervals of $\Delta_c(r)$ that cover zero for $r > 0$. This implies that the posterior predictive intervals for most studies provide no evidence against our model.

2.4 Simulation Studies And Sensitivity Analysis

We briefly discuss our findings of a sensitivity analysis and a study of robustness to model misspecification. Full details are available in Appendix A.

To assess sensitivity to prior specification, we vary several prior and hyperprior values. We investigate nine different prior scenarios. Our conclusion is that the number of population centers is somewhat sensitive to prior specification. This is not surprising as the population centers are latent and partially removed from the data by the second level of our hierarchy. However, our main focus is on the posterior intensity functions, the location and variability of the population centers, and the location and variability of the activation centers about population centers. Specifically, examining the amygdala ROIs, we find that the intensity functions and locations are quite stable, as are the volumes of the 95% credible ellipsoids of the activation centers. However, the volumes of the 95% credible ellipsoids of the (amygdalae) population centers are somewhat sensitive to the prior settings (see Table A.3 in Appendix A).

To assess robustness to model misspecification we simulate data from three different models. The first two data sets are simulated according to our model hierarchy. The first data set is simulated directly from our model. The second data set is simulated from a Matérn cluster process (Møller and Waagepetersen 2004). Our model is resilient to this model misspecification. The third data set is not simulated according to any hierarchy. Foci are drawn directly from a specific intensity function (simulation C, in Appendix A). Here we find that the true activation center process is well approximated by our model. However, in this case, some care is needed in the interpretation of “population centers”.

2.5 Discussion

In this paper, we present a Bayesian hierarchical spatial cluster modeling approach that is novel for neuroimaging meta analysis. Our model provides extra information and results that previously proposed methods cannot; and, as opposed to all current CBMA methods, our model is not massively univariate. With our modeling approach, we can focus attention on specific regions of interest and provide point estimates of the population centers as well estimates of the precision of the population centers and the precision of the activation centers that cluster about each population center. By introducing latent study centers, our model minimizes the potential bias induced by multiple foci per activation region. Our model also accounts for scatter noise (foci that don’t cluster) by modeling them as a homogeneous process. Furthermore, it is a trivial matter to include study weights into our model that account for differences in publication/study fidelity by weighting the variances of the cluster processes.

One potential drawback of our modeling approach is that practitioners may not be used to thinking in terms of spatial models and their related intensity functions. Rather, they are used to the massive univariate approach (voxel by voxel assessment) of current CBMA methods. Nevertheless, our modeling approach does offer the practitioner important information that other methods, to date, cannot provide (see the list in the Introduction).

Future directions within our modeling approach is to incorporate multiple sources of information into study weights such as sample size, nominal significance level, and whether or not the study adjusted for multiple comparisons, to name a few. These various sources could be combined into a single score via principal components analysis and the score discretized by ranking and thresholding on the n-tiles of the first principal component. Another direction would be to account for publication bias in our model. One potential avenue to pursue is to consider the activation centers as a thinning of a marked point process and model the retention probability as a function of, say, the probability of a negative study (that was not published).

Table 2.1: Data Summaries.

Descriptive statistics

	Min.	Median	Mean	Max.
Studies per publication	1	2	2.67	12
Foci per study	1	4	5.67	47
Subjects per publication	4	11	12.23	40

Frequency of modality and inference method

	Fixed	Random	Total
fMRI	32	74	106
PET	48	10	58
Total	80	84	164

Freqency of study emotion type

Emotions	affective	anger	disgust	fear	happy	mixed	sad	surprise	Total
Number	175	26	44	68	36	41	45	2	437

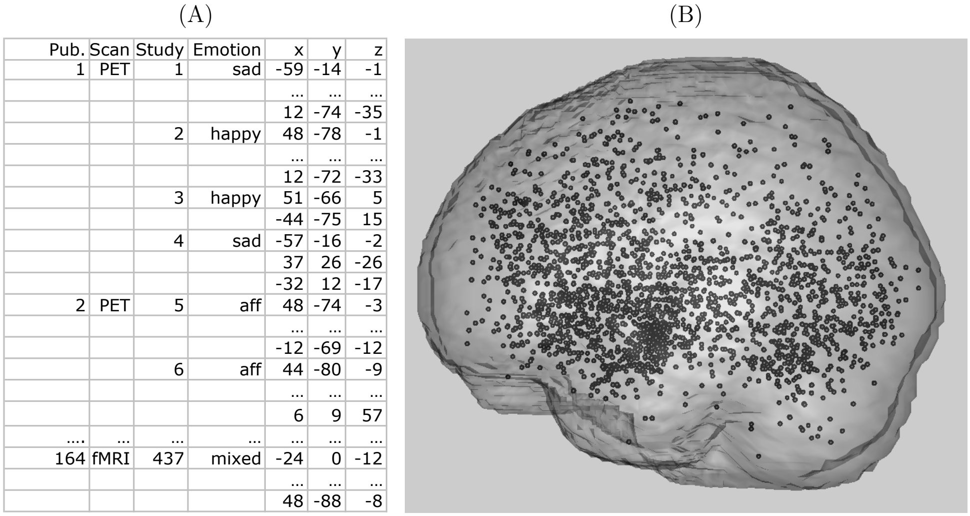
Frequency of corrected and uncorrected thresholds used (intermediate P thresholds rounded up)

P threshold	0.00001	0.0001	0.001	0.005	0.01	0.05	0.1	Missing	Total
Corrected	0	1	12	5	9	89	1	0	117
Uncorrected	1	15	152	48	47	42	0	0	305
Missing	0	0	3	0	0	1	0	11	15
Total	1	16	167	53	56	132	1	11	437

Table 2.2: The 95% credible ellipsoid volumes for population and activation centers.

	Center	Location (mm) x	y	z	95% Credible Ellipsoid Volume (mm^3)
Lt. amygdala	Pop.	-20.3	-5.8	-19.8	138.3
	Act.	-20.7	-6.1	-19.1	12473.8
	modALE	-22.0	-6.0	-16.0	NA
	MKDA	-24.0	-4.0	-18.0	NA
Rt. amygdala	Pop.	23.2	-6.9	-19.7	72.8
	Act.	23.2	-6.2	-19.7	8558.8
	modALE	20.0	-4.0	-18.0	NA
	MKDA	22.0	-4.0	-18.0	NA

Note: Volume of a human brain is about 1,450,000 mm^3
volume of L./R. amygdala is 3,120/3,192 mm^3



(A)

Pub.	Scan	Study	Emotion	x	y	z
1	PET	1	sad	-59	-14	-1
				…	…	…
				12	-74	-35
		2	happy	48	-78	-1
				…	…	…
				12	-72	-33
		3	happy	51	-66	5
				-44	-75	15
		4	sad	-57	-16	-2
				37	26	-26
				-32	12	-17
2	PET	5	aff	48	-74	-3
				…	…	…
				-12	-69	-12
		6	aff	44	-80	-9
				…	…	…
				6	9	57
….	…	…	…	…	…	…
164	fMRI	437	mixed	-24	0	-12
				…	…	…
				48	-88	-8

(B)

Figure 2.1: Panel (A): A subset of the emotion meta analysis data set. Panel (B): All foci, (x, y, z) locations, from all studies plotted in the MNI brain template.

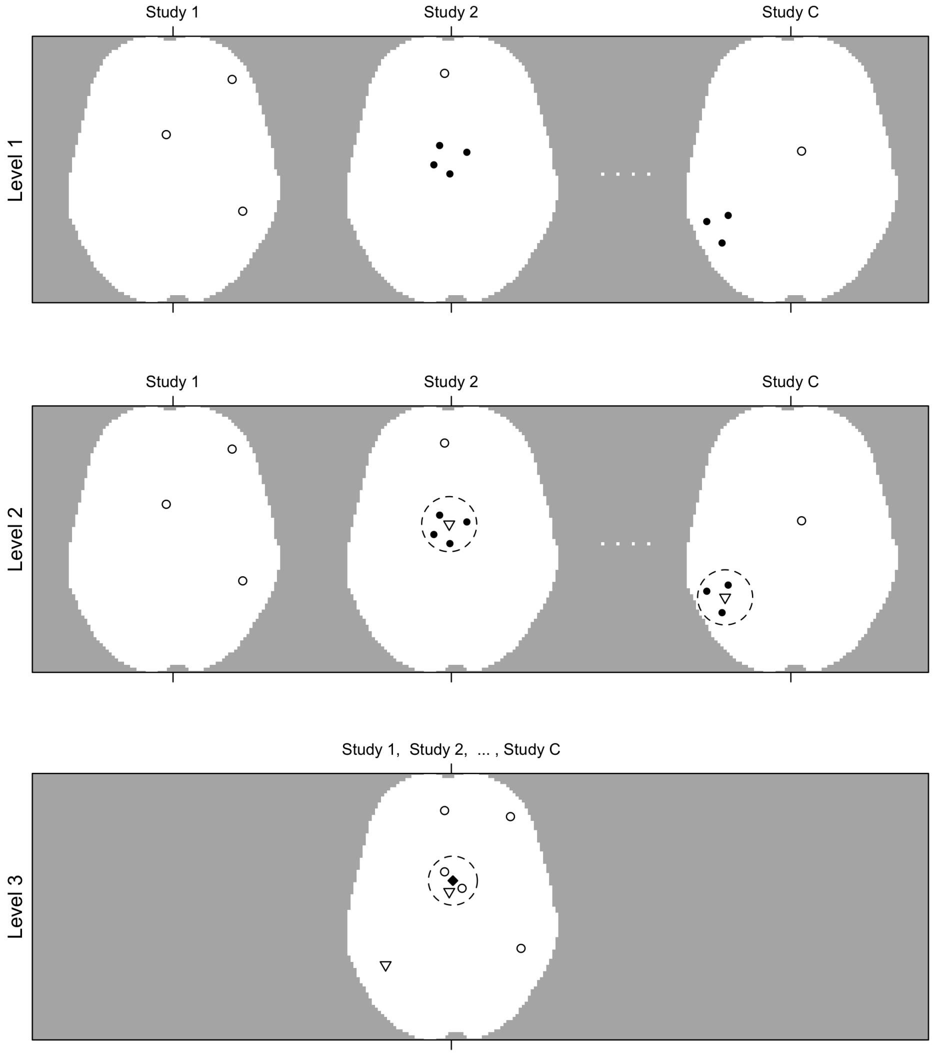


Figure 2.2: Hierarchical Model Illustration: Level 1: The foci from each study, $\mathbf{x}_c$, are the observed data, here shown as open and filled circles. The open circles represent singly reported foci, $\mathbf{x}_c^0$, and the solid circles represent multiply reported foci, $\mathbf{x}_c^1$. Whether foci are multiply or single reported is a latent property, δ_x. Level 2: Multiple reported foci cluster about latent study activation centers, $\mathbf{y}_c$, (open triangles) with the dashed circles representing $\boldsymbol{\Psi}_y$. Singly reported foci in level 1 are shown in level 2 as open circles as they, along with the study activation centers, may cluster together in level 3. Level 3: Activation centers, $\mathbf{x}_c^0$ from level 1 and $\mathbf{y}_c$ from level 2, may cluster about a population center, $\mathbf{z}$ (filled diamond) with the dashed circle representing $\boldsymbol{\Sigma}_z$ or may fail to cluster and are modeled as background scatter and outliers.

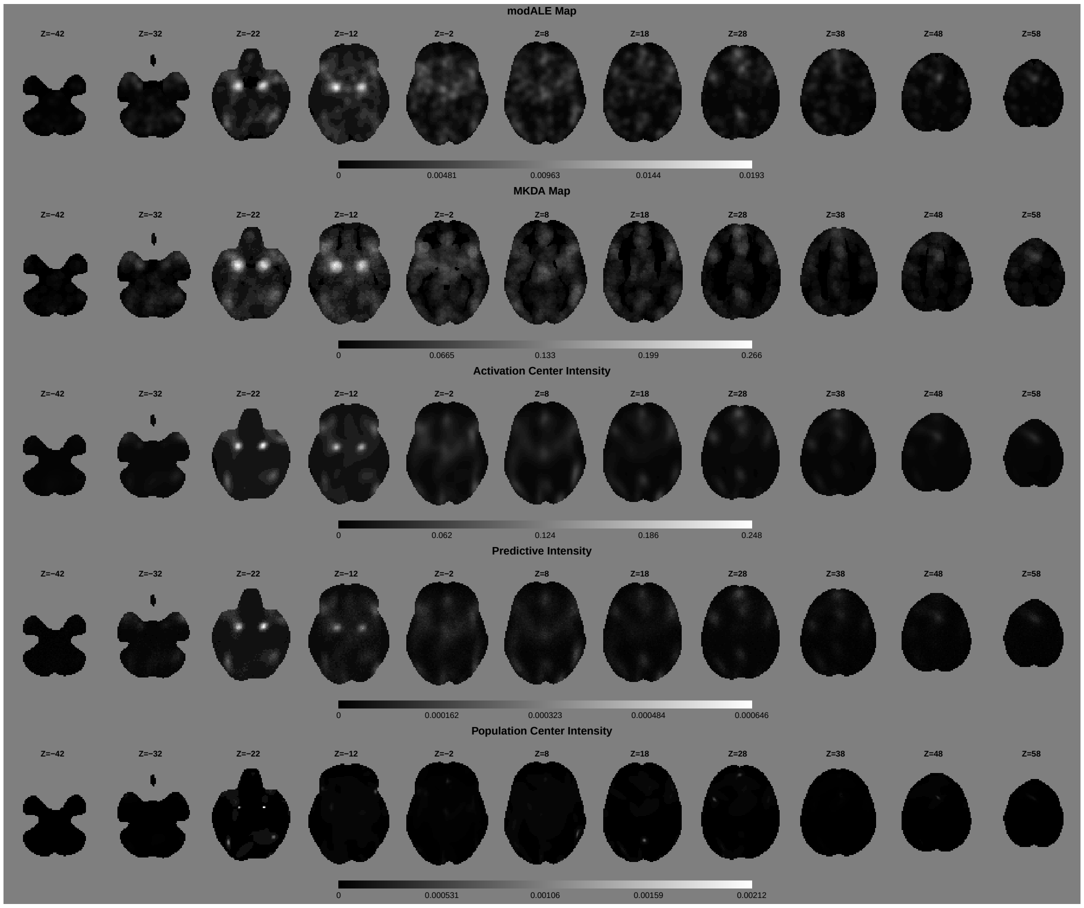
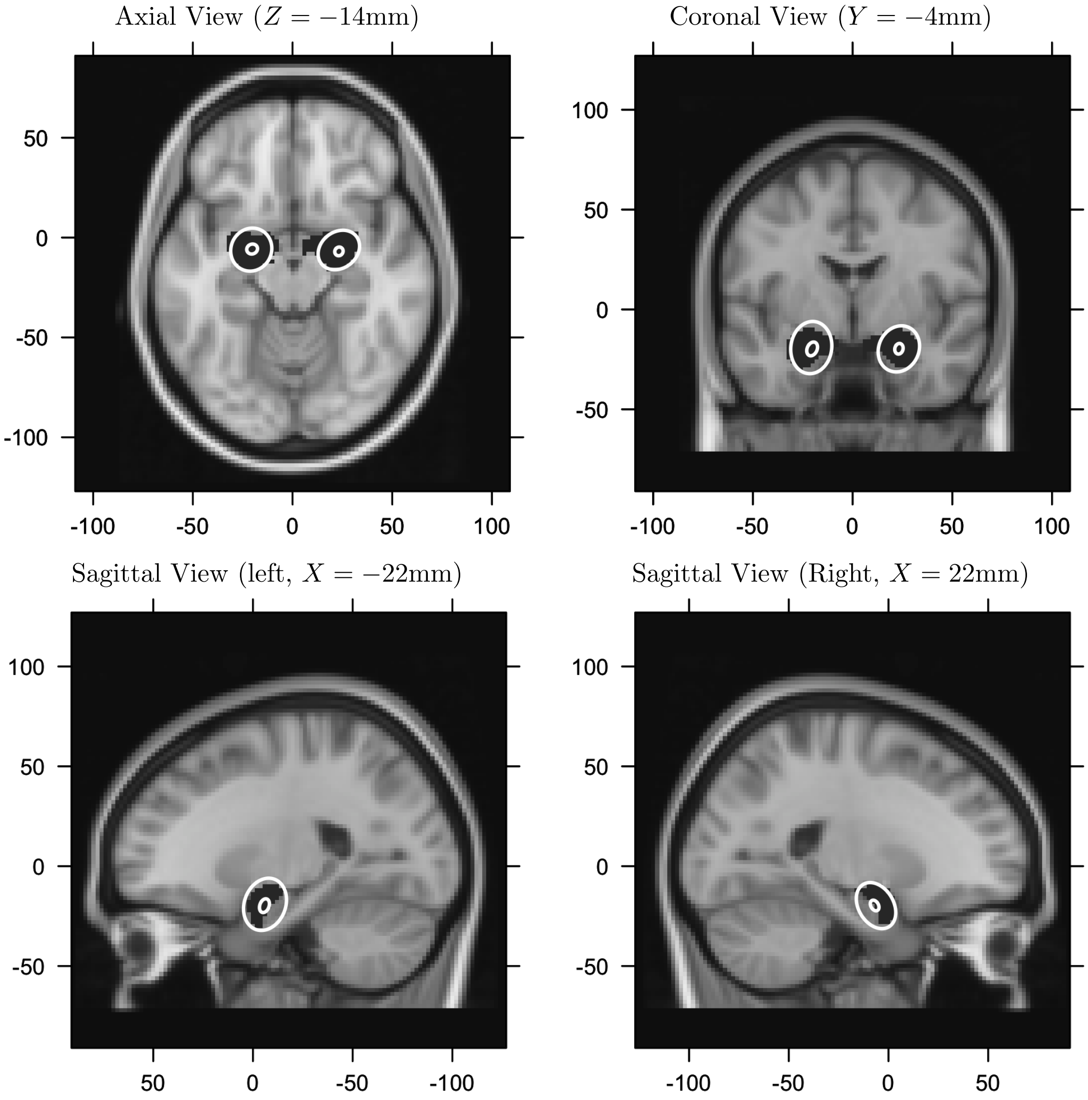


Figure 2.3: Qualitative comparison of the modALE map, MKDA map, the posterior expected activation center intensity function, the posterior predictive intensity for a new study and the posterior expected population center intensity function. We stress here that the gray scale values of the modALE, MKDA and intensity maps are not comparable as their interpretations are not comparable. Qualitatively, the first three rows are similar. The population intensities, however, are much more focused than the activation center intensities, especially in slices $Z = -22, 18, 28$. This reflects the larger variability of activation centers about the population center than the variability of the population centers, themselves.

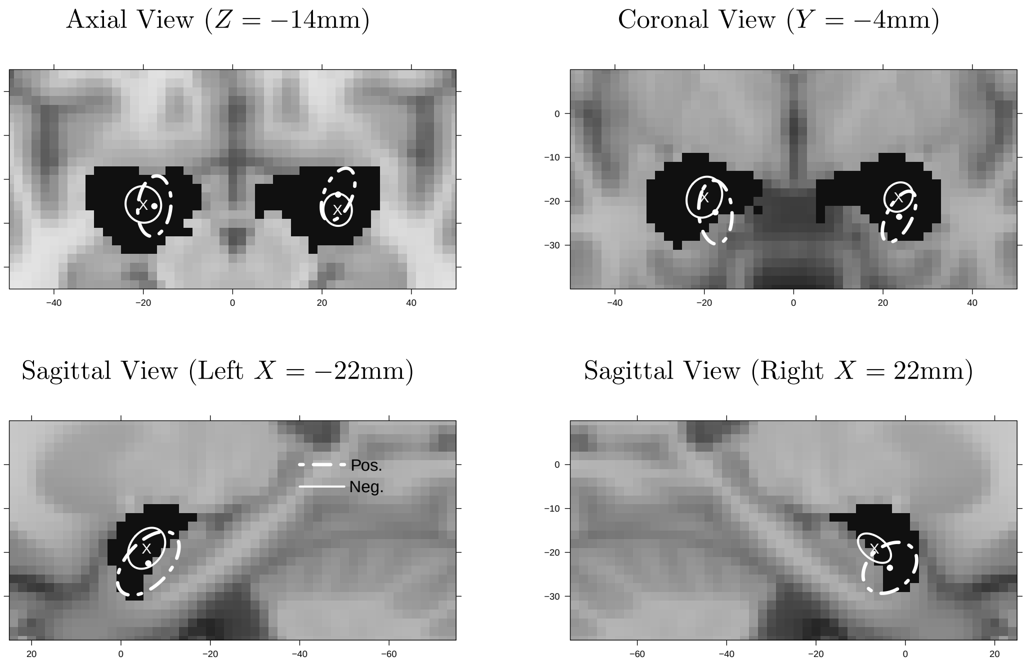

Axial View ($Z = -14$mm)

Coronal View ($Y = -4$mm)

Sagittal View (left, $X = -22$mm)

Sagittal View (Right, $X = 22$mm)

Figure 2.4: The 95% marginal credible ellipses. Large, ellipses are the marginal ellipses of the activation centers. Small, ellipses are the marginal ellipses of the population centers within each amygdala. The black regions (masks), covered by the white ellipses, are the amygdalae.


Axial View ($Z = -14$mm)

Coronal View ($Y = -4$mm)

Sagittal View (Left $X = -22$mm)

Sagittal View (Right $X = 22$mm)

Figure 2.5: The 95% marginal credible ellipses for population centers from positive (dashed ellipses) and negative (solid ellipses) emotion studies. The “x” and the circle represent the centers of the respective ellipses.

CHAPTER III

Hierarchical Poisson/Gamma Random Field Models For Multi-type Spatial Point Patterns

Motivated by the analysis of multiple realizations and groups of spatial point patterns, we propose a non-parametric Bayesian modeling approach that extends the Poisson/Gamma random field model introduced by Wolpert and Ickstadt (1998a). In particular, each group of point patterns is modeled as a Poisson point process driven by a random intensity that is a kernel convolution of gamma random field. The group-level gamma random fields are linked and modeled as a realization of a common gamma random field shared by all groups. We propose an hybrid algorithm with adaptive reject sampling (ARS) embedded in a Markov chain Monte Carlo (MCMC) algorithm for posterior inference. We illustrate our models on simulated examples and two real data sets.

3.1 Introduction

The Poisson/gamma random field (PGRF) model (Wolpert and Ickstadt 1998a) is a nonparametric Bayesian approach to fitting spatial point processes. It has been used in numerous applications (Ickstadt and Wolpert 1999, Best et al. 2000; 2002, Stoyan and Penttinen 2000, Niemi and Fernández 2010, Woodard et al. 2010, Lim and Dass 2011) due to its robustness and computational efficiency. In the PGRF model, the sampling distribution is a Poisson point process. The intensity function, in turn, is modeled as a convolution of a spatial kernel and a gamma random field.

The PGRF model assumes a single realization of a spatial point process. However, there is growing interest in analyzing multiple spatial point patterns (Baddeley 2010) motivated by problems in the biological and medical sciences such as a disease mapping (Lawson et al. 2000, Diggle 1990, Diggle and Rowlingson 1994, Baddeley et al. 2005); studies in spatial epidemiology, where each point represents either a case or a control; the analyses of plant ecologies (Brix and Møller 2001, Diggle 1981; 1983, Diggle and Milne 1985, van Lieshout and Baddeley 1999), where the locations of plants are marked by the species classification of each plant; and the meta-analysis of functional neuroimaging studies (Fox et al. 1997, Kober et al. 2008, Eickhoff et al. 2009, Wager et al. 2009, Kang et al. 2011), which combines foci, (x, y, z) coordinates in a three dimensional brain, reported from many independently performed studies.

In general, such data can be represented by two mathematically equivalent formulations that lead to different statistical approaches. The first is to treat the data as a single pattern of n points where each point location x_i is marked by its type z_i. This results in a marked point pattern, or more specifically, a multi-type point pattern when type z_i admits only integer values. Statistical methods for this type of data analyze the marks and locations jointly or conditionally (Diggle 1990, Diggle and Rowlingson 1994, Höogmander and Särkkä 1999). Alternatively, the data can be represented as a multivariate observation $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_J)$ where $\mathbf{x}_j$ is the point pattern of type j.

Statistical methods to analyze multivariate point patterns include Diggle and Milne (1985), van Lieshout and Baddeley (1999), Diggle et al. (2005). In particular, Diggle and Milne (1985) investigated statistical approaches for bivariate point pattern by fitting bivariate Cox processes. van Lieshout and Baddeley (1999) proposed several summary statistics to quantify the dependence between the different types of point patterns. Diggle et al. (2005) addressed the spatial segregation in a multivariate point process via nonparametric estimation of ratios of component-wise intensities. Multivariate log Gaussian Cox processes have also been extensively studied (Møller et al. 1998, Brix and Møller 2001, Møller and Waagepetersen 2004; 2007). where the multivariate point pattern $\mathbf{x}$ is assumed to be a realization of a multivariate Cox process driven by $\mathbf{z} = (\exp(\mathbf{y}_1), \ldots, \exp(\mathbf{y}_J))$ where $\mathbf{y} = (\mathbf{y}_1, \ldots, \mathbf{y}_J)$ is a J-dimensional Gaussian random field.

In this article, we propose an alternative approach to fitting multivariate spatial point patterns by extending the PGRF model to an hierarchical PGRF (HPGRF) model, in the same spirit as Teh et al. (2006) extend Dirichlet processes to hierarchical Dirichlet processes. Specifically, we consider each type of spatial point pattern as a realization of one PGRF where the gamma random fields across different types are assumed to be drawn from a common gamma random field. This implies that the random intensity functions for different types are linked which allows sharing of clusters between different types in the multivariate spatial point pattern. An illustrative example is the analysis of spatial locations of two cohorts of plants: newly emergent and 1-year-old bramble canes in a 9 meter squared rectangular plot (Hutchings 1979). Diggle and Milne (1985) analyzed this data using a bivariate linked Cox process. In their model, spatial correlation between the two types is induced by assuming that the random intensity functions for the types are equal up to a constant. Thus, their is a strong restriction that the spatial clustering patterns of the two types are the same. Relaxing this assumption, the HPGRF model allows each type to have its own intensity while allowing common clustering patterns between the two types. We discuss this example in more detail in Section 3.3.1.

As noted by Wolpert and Ickstadt (1998a), the PGRF model is analogous to the mixture of Dirichlet processes (MDP) model for random distributions of non-point pattern data (Antoniak 1974, Ferguson 1973, Escobar and West 1995). There is growing interest in extending the MDP model to accommodate dependence between the distributions (MacEachern 1999, Teh et al. 2006, Rodriguez et al. 2008), where the hierarchical DP (HDP) by Teh et al. (2006) has been widely applied to many problems (Xing and Sohn 2007, Hoffman et al. 2008, Blei et al. 2010, Fox et al. 2011). The HDP model assumes that the multiple group specific distributions are drawn from different DPs with a common baseline measure which is in turn a draw from another DP. Our HPGRF model generalizes the PGRF in a similar manner, allowing the different intensity/probability measures to share the same set of atoms but have distinct sets of weights.

The main motivating example comes from a functional neuroimaging meta analysis of emotions (Kober et al. 2008) consisting of a set of functional neuroimaging studies of emotion. Each study reports a list of three-dimensional foci: that is, the peak activation locations in statistically significant regions of activation within the brain. A primary goal of functional neuroimaging meta-analysis studies is to find clusters of foci across all studies. Clusters appear in regions with high foci intensity. That is, where a preponderance of studies report activation. We assume, with reason, that the foci reported from different studies comprise multiple realizations of a spatial point process driven by a latent intensity function that identifies the consistent activation, if any exists, across studies. Thus, the HPGRF model can be used to estimate the common intensity function across studies. This meta-analysis data set also contains information about the types of emotions studied. Thus, another primary goal is to identify foci clustering across the different types of emotion. For example, the amygdalae are thought to play an important role in the processing of emotions (Phelps and LeDoux 2005). This example is discussed in detail in Section 3.3.2.

The remainder of this chapter is organized as follows. Section 3.2 presents the HPGRF model for multi-type spatial point patterns. We first describe the model for a single realization in Section 3.2.1, where we discuss the model properties, a data augmentation scheme and a Lévy measure construction for the model representation, and a hybrid algorithm with adaptive reject sampling (ARS) embedded in a Markov chain Monte Carlo (MCMC) algorithm to simulate the posterior. Then we extend the model for multiple realizations in Section 3.2.2. We illustrate our proposed models on two motivating examples in Section 3.3 and assess model performance via simulation studies in Section 3.4. Finally, we draw conclusions with a brief discussion in Section 3.5.

3.2 The Model

In this section, we first extend the PGRF model to a hierarchical PGRF model for multi-type spatial point processes. Then we motivate and further extend the HPGRF model to fit multiple independent realizations of the spatial point process for each type. In this article, all the spatial point patterns are defined on a region $\mathcal{B}$ which is a subregion of the d dimensional Euclidean space, $\mathbb{R}^d$.

3.2.1 Hierarchical Poisson/Gamma Random Fields

Let J denote the number of distinct types of spatial point patterns. Let $\mathbf{y}_j$, $j = 1, \ldots, J$, denote the observed point pattern for type j on region $\mathcal{B}$. We assume that $\mathbf{y}_j$ is a single realization of a Poisson point process $\mathbf{Y}_j$ with intensity measure

$\Lambda_j(dy)$ which, in turn, is a convolution of a random measure $G_j(d\mathrm{x})$ with a kernel measure $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x})$, where σ_j^2 is the kernel variance. For each j, $G_j(d\mathrm{x})$ is assumed to be a gamma random field with base measure $G_0(d\mathrm{x})$ and scale parameter τ, and the $G_j(d\mathrm{x})$ are mutually independent. To introduce dependence between different types of spatial point patterns, let $G_0(d\mathrm{x})$ be a gamma random field with base measure $\alpha(d\mathrm{x})$ and scale parameter β. In summary, we have

$$\begin{aligned}
[\mathbf{Y}_j \mid \Lambda_j(d\mathrm{y})] &\sim \mathcal{PP}\{\mathcal{B}, \Lambda_j(d\mathrm{y})\}, \quad \Lambda_j(dy) = \int_{\mathcal{B}} K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) G_j(d\mathrm{x}), \\
[G_j(d\mathrm{x}) \mid G_0(d\mathrm{x}), \tau] &\overset{i.i.d.}{\sim} \mathcal{GRF}\{G_0(d\mathrm{x}), \tau\}, \\
[G_0(d\mathrm{x}) \mid \alpha(d\mathrm{x}), \beta] &\sim \mathcal{GRF}\{\alpha(d\mathrm{x}), \beta\}.
\end{aligned} \tag{3.1}$$

Here, the notation $\mathcal{PP}\{A, \Lambda(d\mathrm{x})\}$ denotes a Poisson point process on region A, a subset of $\mathbb{R}^d$, with intensity measure $\Lambda(d\mathrm{x})$, and $\mathcal{GRF}\{a(d\mathrm{x}), b\}$ represents a gamma random field with base measure $a(d\mathrm{x})$ and scale parameter b. Suppose $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x})$ and $\Lambda_j(dy)$ are both dominated by a reference measure $\Pi(dy)$. In this article, $\Pi(dy)$ is the Lebesgue measure. Furthermore, we have $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) = \int_{\mathcal{B}} k_{\sigma_j^2}(\mathrm{y}, \mathrm{x}) \Pi(dy)$ and $\Lambda_j(dy) = \int_{\mathcal{B}} \lambda_j(\mathrm{y}) \Pi(dy)$, where $k_{\sigma_j^2}(\mathrm{y}, \mathrm{x})$ is a kernel function and $\lambda_j(\mathrm{y})$ is an intensity function. The intensity measure $\Lambda_j(dy)$ is the individual level intensity for each point process $\mathbf{Y}_j$. We define the "population mean" intensity as

$$\Lambda_0(dy) = \frac{1}{\tau} \int_{\mathcal{B}} \tilde{K}(d\mathrm{y}, \mathrm{x}) G_0(d\mathrm{x}), \tag{3.2}$$

where $\tilde{K}(d\mathrm{y}, \mathrm{x}) = J^{-1} \sum_{j=1}^{J} K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x})$. This implies that

$$\Lambda_0(dy) = \frac{1}{J} \sum_{j=1}^{J} E[\Lambda_j(dy) \mid G_0, \sigma_j^2, \tau],$$

the average of the expected intensity measures over different types. The HPGRF is similar to the HDP model by Teh et al. (2006) and induces spatial correlation between the number of points in any two regions of interest between types. More specifically, the following model properties hold.

Model Properties

Let $\delta_z(A)$ denote the Dirac delta measure, i.e. $\delta_z(A) = 1$ if $z \in A$, $\delta_z(A) = 0$, otherwise. For any $A \subseteq \mathcal{B}$ and a generic point process $\mathbf{Z}$, define the counting measure $N_{\mathbf{Z}}(A) = \sum_{\mathrm{z} \in \mathbf{Z}} \delta_{\mathrm{z}}(A)$, that counts the number of points of $\mathbf{Z}$ in A. The conditional mean and covariance structure of $N_{\mathbf{Y}_j}(A)$ for the HPGRF model are summarized in the following theorem whose proof is given in Appendix B.

Theorem 1. *Within type j, for any $A, B \subseteq \mathcal{B}$,*

$$\mathrm{E}\{N_{\mathbf{Y}_j}(A) \mid \sigma_j^2, \tau, \alpha, \beta\} = \frac{1}{\tau\beta} \int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) \alpha(d\mathrm{x}). \tag{3.3}$$

$$\begin{aligned} &\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_j}(B) \mid \sigma_j^2, \tau, \alpha, \beta\} \\ &\quad = \frac{1}{\tau\beta} \int_{\mathcal{B}} K_{\sigma_j^2}(A \cap B, \mathrm{x}) \alpha(d\mathrm{x}) + \frac{1+\beta}{\tau^2\beta^2} \int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) K_{\sigma_j^2}(B, \mathrm{x}) \alpha(d\mathrm{x}). \end{aligned} \tag{3.4}$$

Between types j and k ($j \neq k$),

$$\begin{aligned} &\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_k}(B) \mid \sigma_j^2, \sigma_k^2, \tau, \alpha, \beta\} = \\ &\frac{1}{\tau^2\beta^2} \int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) K_{\sigma_k^2}(B, \mathrm{x}) \alpha(d\mathrm{x}). \end{aligned} \tag{3.5}$$

This theorem shows that as an a priori property, the number of points in A and B are correlated for different types of point processes, whether or not A and B are disjoint. When $\sigma_j^2 = \sigma_k^2$, $j \neq k$, (3.4) and (3.5) show that the covariance within the same group is larger than the covariance from different types. Posterior inference of the HPGRF model is realized by the following model representation.

Augmentation and Complete Data Model

Wolpert and Ickstadt (1998a) propose an alternative model representation based on data augmentation that produces a more efficient MCMC algorithm for posterior estimation of the PGRF model. Modifying their approach, we attach a mark to each point in $\mathbf{Y}_j$. Based on the model (3.1), $N_{\mathbf{Y}_j}(\mathcal{B})$ is a Poisson random variable with mean $\Lambda_j(\mathcal{B})$. Given $N_{\mathbf{Y}_j}(\mathcal{B})$, $G_j(d\mathrm{y})$ and σ_j^2, all points in $\mathbf{Y}_j$ are mutually independent and follow the same distribution determined by $\Lambda_j(d\mathrm{y})$: for each point $Y_j \in \mathbf{Y}_j$,

$$[Y_j \mid N_{\mathbf{Y}_j}(\mathcal{B}), G_j(d\mathrm{y}), \sigma_j^2] \sim \frac{\Lambda_j(d\mathrm{y})}{\Lambda_j(\mathcal{B})} = \frac{\int_{\mathcal{B}} K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) G_j(d\mathrm{x})}{\Lambda_j(\mathcal{B})}. \tag{3.6}$$

For each $Y_j \in \mathbf{Y}_j$, we resolve this mixture by drawing an auxiliary random variable $X_j = \mathrm{x} \in \mathcal{B}$ from the distribution,

$$[X_j \mid Y_j, N_{\mathbf{Y}_j}(\mathcal{B}), G_j(d\mathrm{x}), \sigma_j^2] \sim \frac{k_{\sigma_j^2}(Y_j, \mathrm{x}) G_j(d\mathrm{x})}{\lambda_j(Y_j)}. \tag{3.7}$$

Define $(\mathbf{X}_j, \mathbf{Y}_j) = \{(X_j, Y_j), Y_j \in \mathbf{Y}_j\}$. Then we can show that $(\mathbf{X}_j, \mathbf{Y}_j)$ is a Poisson point process on $\mathcal{B} \times \mathcal{B}$ with intensity measure $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) G_j(d\mathrm{x})$:

$$[(\mathbf{X}_j, \mathbf{Y}_j) \mid K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) G_j(d\mathrm{x})] \quad \sim \quad \mathcal{PP}\{\mathcal{B}, K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) G_j(d\mathrm{x})\}, \tag{3.8}$$

where we only observe the realization of $\mathbf{Y}_j$, while $\mathbf{X}_j$ is the collection of latent marks. By integrating out $\mathbf{X}_j$, the marginal distribution of $\mathbf{Y}_j$ reduces to model (3.1).

The Lévy Measure Construction

Several methods to simulate gamma random fields have been proposed including Bondesson (1982), Damien et al. (1995) and Wolpert and Ickstadt (1998b). The inverse Lévy measure algorithm (Wolpert and Ickstadt 1998a;b) provides an efficient approach and has been successfully applied to the PGRF model. We represent the algorithm in the following theorem.

Theorem 2. *Let* $\theta_m \overset{i.i.d.}{\sim} \tilde{\alpha}(d\mathrm{x}) = \alpha(d\mathrm{x})/\alpha(\mathcal{B})$*,* $\nu_m = E_1^{-1}\{\zeta_m/\alpha(\mathcal{B})\}/\beta$*, and* $\zeta_m = \sum_{l=1}^{m} e_l$*, for* $m = 1, 2, \ldots$*, where* $e_l \overset{i.i.d.}{\sim} Exp(1)$*, i.e. the standard exponential distribution, and* $E_1(t) = \int_t^\infty e^{-u}u^{-1}du$*. Let* $\Gamma[d\mathrm{x}] = \sum_{m=1}^{\infty} \nu_m \delta_{\theta_m}(d\mathrm{x})$*, then*

$$\Gamma(d\mathrm{x}) \sim \mathcal{GRF}\{\alpha(d\mathrm{x}), \beta\}.$$

Note that $\tilde{\alpha}(d\mathrm{x})$ is a probability measure generated by the normalization of $\alpha(d\mathrm{x})$. The sequence $\{\zeta_m\}_{m=1}^{M}$ denotes the jumps of the standard Poisson process on $\mathbb{R}^+$. This is a specific case of Theorem 1 by Wolpert and Ickstadt (1998a). It not only provides an efficient approach to the simulation of a Gamma random field, but it also can be used for model representation to simplify the posterior computation. Let InvLévy$[\alpha(d\mathrm{x}), \beta]$ represent the sampling distribution of $\{(\theta_m, \nu_m)\}_{m=1}^{M}$, generated according to Theorem 2 given the base measure $\alpha(d\mathrm{x})$ and scale parameter β, where M is an extended natural number ($\bar{\mathbb{N}}$). Theorem 2 suggests an alternative representation of model (3.1):

$$G_0(d\mathrm{x}) \; = \; \sum_{m=1}^{\infty} \nu_m \delta_{\theta_m}(d\mathrm{x}), \tag{3.9}$$

where $\{(\theta_m, \nu_m)\}_{m=1}^{\infty} \sim$ InvLévy$\{\alpha(d\mathrm{x}), \beta\}$. Note that G_0 has support at $\{\theta_m\}_{m=1}^{\infty}$. This implies that each G_j necessarily has the same support. Thus, there exist positive random numbers $\mu_{j,m}, j = 1, \ldots, J, m \in \mathbb{N}$, such that

$$G_j(d\mathrm{x}) \; = \; \sum_{m=1}^{\infty} \mu_{j,m} \delta_{\theta_m}(d\mathrm{x}). \tag{3.10}$$

Let $(B_1, \ldots, B_r)$ be any partition of $\mathcal{B}$. Let $A_l = \{m : \theta_m \in B_l\}$ for $l = 1, \ldots, r$. Then $(A_1, \ldots, A_r)$ is a finite partition of the natural numbers. For each j and l, we have $G_j(B_l) \sim Gamma(G_0(B_l), \tau)$ so that $\sum_{m \in A_l} \mu_{j,m} \sim Gamma\left(\sum_{m \in A_l} \nu_m, \tau\right)$. Thus, for $m \in \mathbb{N}$,

$$\mu_{j,m} \sim Gamma(\nu_m, \tau). \tag{3.11}$$

Finally, combining equations (3.8), (3.9), (3.10) and (3.11) we have the following equivalent representation of the HPGRF model:

$$\begin{aligned} \left[(\mathbf{X}_j, \mathbf{Y}_j) \mid \{(\mu_{j,m}, \theta_m)\}_{m=1}^{\infty}, \sigma_j^2\right] &\sim \mathcal{PP}\left\{\mathcal{B}, K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) \sum_{m=1}^{\infty} \mu_{j,m}\delta_{\theta_m}(d\mathrm{x})\right\}, \\ [\mu_{j,m} \mid \nu_m, \tau] &\overset{i.i.d.}{\sim} Gamma\,(\nu_m, \tau)\,, \\ \{(\theta_m, \nu_m)\}_{m=1}^{\infty} &\sim \text{InvLévy}\{\alpha(d\mathrm{x}), \beta\}. \end{aligned} \tag{3.12}$$

In practice, we cannot sample $\{(\mathbf{X}_j, \mathbf{Y}_j)\}_{j=1}^{J}$ according to (3.12), since it requires simulating an infinite number of parameters, which, in fact, reflects the nonparametric nature of the HPGRF model. We suggest the following truncated model: for any finite integer M,

$$\begin{aligned} \left[(\mathbf{X}_j, \mathbf{Y}_j) \mid \{(\mu_{j,m}, \theta_m)\}_{m=1}^{M}, \sigma_j^2\right] &\sim \mathcal{PP}\left\{\mathcal{B}, K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) \sum_{m=1}^{M} \mu_{j,m}\delta_{\theta_m}(d\mathrm{x})\right\}, \\ [\mu_{j,m} \mid \nu_m, \tau] &\overset{i.i.d.}{\sim} Gamma\,(\nu_m, \tau)\,, \\ \{(\theta_m, \nu_m)\}_{m=1}^{M} &\sim \text{InvLévy}\{\alpha(d\mathrm{x}), \beta\}. \end{aligned} \tag{3.13}$$

The following theorem states that model (3.12) can be approximated by the truncated model (3.13) to any desired level of accuracy.

Theorem 3. *For $j = 1, \ldots, J$, for any $\epsilon > 0$ and for any region $A \subseteq \mathcal{B}$, there exists a natural number M_ϵ, such that*

$$E\{\Lambda_j(A) - \Lambda_j^{M_\epsilon}(A) \mid \beta, \tau\} < \epsilon,$$

where $\Lambda_j^M(A) = \sum_{m=1}^{M} \mu_{jm} K_{\sigma_j^2}(A, \theta_m)$ is the conditional expectation of $N_{\mathbf{Y}_j}(A)$ in the truncated model (3.13)*.*

Posterior Computation

Theorem 3 implies that, for a sufficiently large integer M the truncated model (3.13) can well approximate the HPGRF model (3.1). Furthermore, the approximation only involves a fixed number of parameters allowing the computation of the posterior.

The target distribution

Let $(\mathbf{x}_j, \mathbf{y}_j)$ denote a realization of $(\mathbf{X}_j, \mathbf{Y}_j)$, m_j be the observed number of points in $(\mathbf{x}_j, \mathbf{y}_j)$, and $(\mathrm{x}_{j,l}, \mathrm{y}_{j,l}) \in (\mathbf{x}_j, \mathbf{y}_j)$ denote an observed point indexed by l, for $l = 1, \ldots, m_j$. Write $(\mathbf{x}_j, \mathbf{y}_j) = \{(\mathrm{x}_{j,l}, \mathrm{y}_{j,l})\}_{l=1}^{m_j}$, $\boldsymbol{\theta} = \{\theta_m\}_{m=1}^{M}$, $\boldsymbol{\nu} = \{\nu_m\}_{m=1}^{M}$, and $\boldsymbol{\mu}_j = \{\mu_{j,m}\}_{m=1}^{M}$. The joint density of $\{\mathbf{x}_j\}_{j=1}^{J}$, $\{\boldsymbol{\mu}_j\}_{j=1}^{J}$, $\{\sigma_j^2\}_{j=1}^{J}$, $\boldsymbol{\nu}$, $\boldsymbol{\theta}$, τ and β given $\{\mathbf{y}_j\}_{j=1}^{J}$ is proportional to

$$
\begin{aligned}
&\prod_{j=1}^{J} \pi(\mathbf{x}_j, \mathbf{y}_j \mid \boldsymbol{\mu}_j, \boldsymbol{\theta}, \sigma_j^2) \pi(\sigma_j^2) \pi(\boldsymbol{\mu}_j \mid \boldsymbol{\nu}, \tau) \times \pi(\tau) \pi(\boldsymbol{\nu} \mid \beta) \pi(\beta) \pi(\boldsymbol{\theta}) \\
&\quad \propto \prod_{j=1}^{J} \exp\left\{ -\sum_{m=1}^{M} K_{\sigma_j^2}(\mathcal{B}, \theta_m) \mu_{j,m} \right\} \prod_{l=1}^{m_j} \left\{ k_{\sigma_j^2}(\mathrm{y}_{j,l}, \mathrm{x}_{j,l}) \sum_{m=1}^{M} \mu_{j,m} I_{\theta_m}(\mathrm{x}_{j,l}) \right\} \\
&\qquad \times \prod_{j=1}^{J} \left[\pi(\sigma_j^2) \prod_{m=1}^{M} \left\{ \frac{\tau^{\nu_m} \mu_{j,m}^{\nu_m - 1}}{\Gamma(\nu_m)} \exp\{-\tau \mu_{j,m}\} \right\} \right] \\
&\qquad \times \pi(\tau) \exp\{-E_1(\beta \nu_M)\} \prod_{m=1}^{M} \{\nu_m^{-1} \exp\{-\nu_m \beta\}\} \pi(\beta),
\end{aligned}
\tag{3.14}
$$

where $\pi(\mathbf{x}_j, \mathbf{y}_j \mid \boldsymbol{\mu}_j, \boldsymbol{\theta}, \sigma_j^2)$ is the density of $(\mathbf{X}_j, \mathbf{Y}_j)$ with respect to a unit rate Poisson process (Møller and Waagepetersen 2004), $\pi(\boldsymbol{\theta})$ is the density of $\boldsymbol{\theta}$ with respect to $\prod_{m=1}^{M} \tilde{\alpha}(d\mathrm{x})$ and, in this article, $\alpha(d\mathrm{x})$, is assumed to be non-atomic. Then, the θ_m, for $m = 1, 2, \ldots, M$, are distinct with probability one. The densities of other parameters are all with respect to a product of Lebesgue measures. Note that $\pi(\boldsymbol{\theta}) = 1$ in our model. Also, the function $I_x(y)$ is an indicator function with $I_x(y) = 1$ if $x = y$, $I_x(y) = 0$, otherwise. Note that $I_x(y) = I_y(x)$.

Next, we summarize the key steps in the posterior simulation.

- **Sampling $\mathbf{x}_j$**

From (3.14), it is straightforward to obtain the conditional distribution of $\mathrm{x}_{j,l}$

given all other parameters:

$$\Pr(\mathrm{x}_{j,l} = \theta_m \mid \cdot) \propto \mu_{j,m} k_{\sigma_j^2}(\mathrm{y}_{j,l}, \theta_m). \tag{3.15}$$

- **Sampling $\boldsymbol{\theta}$**

The conditional density of $\boldsymbol{\theta}$ with respect to $\prod_{m=1}^{M} \tilde{\alpha}(d\mathrm{x})$ is

$$\begin{aligned} \pi(\boldsymbol{\theta} \mid \cdot) \propto \exp &\left\{ - \sum_{m=1}^{M} \left[\sum_{j=1}^{J} K_{\sigma_j^2}(\mathcal{B}, \theta_m) \mu_{j,m} \right] \right\} \\ &\times \prod_{j=1}^{J} \prod_{l=1}^{m_j} \left[\sum_{m=1}^{M} \mu_{j,m} I_{\mathrm{x}_{j,l}}(\theta_m) \right]. \end{aligned} \tag{3.16}$$

This implies that for all j, l, $\sum_{m=1}^{M} I_{\mathrm{x}_{j,l}}(\theta_m) > 0$. There exists at least one θ_m that is equal to $\mathrm{x}_{j,l}$. Suppose there are $\tilde{N}$ distinct points among $\{\mathrm{x}_{j,l}\}$, denoted as $\{s_1, \ldots, s_{\tilde{N}}\}$. Due to the symmetry of $\{\mu_{1,m}, \ldots, \mu_{J,m}, \theta_m\}_{m=1}^{M}$ in (3.16) and noting that $\theta_1, \cdots, \theta_m$ are distinct points, each θ_m must be equal to one of $s_1, \ldots, s_{\tilde{N}}$. Thus, to sample $\boldsymbol{\theta}$, we first draw a random permutation $\{p_1, \ldots, p_m\}$ of $\{1, \ldots, M\}$, then set $\theta_{p_m} = s_m$ for $1 \leq m \leq \tilde{N}$, and for $m > \tilde{N}$, we draw θ_{p_m} according to the following conditional density with respect to $\alpha(d\mathrm{x})$,

$$\pi(\theta_{p_m} \mid \cdot) \propto \exp \left\{ - \sum_{j=1}^{J} K_{\sigma_j^2}(\mathcal{B}, \theta_{p_m}) \mu_{j,p_m} \right\}. \tag{3.17}$$

- **Sampling $\boldsymbol{\mu}_j$**

Theorem 4. *The full conditional distribution of $\mu_{j,m}$, for $j = 1, \ldots, J$ and $m = 1, \ldots, M$, is given by*

$$[\mu_{j,m} \mid \cdot] \sim \mathrm{Gamma}[\nu_m + b_{jm}, K_{\sigma_j^2}(\mathcal{B}, \theta_m) + \tau]. \tag{3.18}$$

where $b_{jm} = \sum_{l=1}^{m_j} I_{\theta_m}(\mathrm{x}_{j,l})$.

Proof in given in Appendix B.

- **Sampling $\boldsymbol{\nu}$**

Theorem 5. *The conditional distribution of ν_m, given all other parameters, is*

$$(3.19)\ \pi(\nu_m \mid \cdot) \propto \begin{cases} \frac{c_m^{\nu_m}}{\Gamma(\nu_m)}\nu_m^{-1}, & m = 1, \dots, M-1 \\ \exp\{-E_1(\beta\nu_M)\}\frac{c_M^{\nu_M}}{\Gamma(\nu_M)}\nu_M^{-1}, & m = M \end{cases}$$

where $c_m = \tau^J \prod_{j=1}^J \mu_{j,m} e^{-\beta}$ and $\pi(\nu_m \mid \cdot)$, for $m = 1, \dots, M$, are log concave functions.

Proof in given in Appendix B. Theorem 5 implies that we can draw ν_m for $m = 1, \dots, M$ from (3.19) by using the adaptive rejection sampling (ARS) algorithm (Gilks and Wild 1992).

Sampling hyperparameters

We update σ_j^2, for $j = 1, \dots, J$, τ using Metropolis within Gibbs sampling. In this article, we choose $k_{\sigma_j^2}(\mathrm{x}, \mathrm{y}) = \left(\frac{1}{2\pi\sigma_j^2}\right)^{d/2} \exp\{-\|\mathrm{x} - \mathrm{y}\|^2/(2\sigma_j^2)\}$ and assume, a priori, $\sigma_j^{-2} \sim \text{Uniform}[a_\sigma, b_\sigma]$, $\tau \sim \mathit{Gamma}(a_\tau, b_\tau)$ and $\beta \sim \mathit{Gamma}(a_\beta, b_\beta)$. The full conditional of σ_j^2 is

$$\begin{aligned} &\pi(\sigma_j^2 \mid \cdot) \\ (3.20) \qquad &\propto \exp\left[-\sum_{m=1}^M \left\{K_{\sigma_j^2}(\mathcal{B}, \theta_m)\mu_{j,m}\right\} - \frac{S_j^2}{\sigma_j^2} - \frac{m_j d}{2}\log(\sigma_j^2)\right] I_{[a_\sigma, b_\sigma]}(\sigma_j^{-2}), \end{aligned}$$

where $S_j^2 = \frac{1}{2}\sum_{l=1}^{m_j} \|\mathrm{x}_{j,l} - \mathrm{y}_{j,l}\|^2$. Thus, to update σ_j^2, we use a random walk and first draw $\sigma_j^{2*} \sim N(\sigma_j^2, \theta_\sigma^2)$, if $\sigma_j^{2*} \in (a_\sigma, b_\sigma)$, then set $\sigma_j^2 = \sigma_j^{2*}$ with probability $\min\{r(\sigma), 1\}$, where $r(\sigma)$ is

$$\exp\left[\sum_{m=1}^M \left\{\left[K_{\sigma_j^2}(\mathcal{B}, \theta_m) - K_{\sigma_j^{2*}}(\mathcal{B}, \theta_m)\right]\mu_{j,m}\right\} + S_j^2\left(\frac{1}{\sigma_j^2} - \frac{1}{\sigma_j^{2*}}\right)\right]\left(\frac{\sigma_j^2}{\sigma_j^{2*}}\right)^{\frac{1}{2}m_j d}$$

The full conditional of τ is

$$(3.21) \qquad \pi(\tau \mid \cdot) \propto \tau^{J\sum_{m=1}^M \nu_m + a_\tau - 1} \exp\left\{-\left(b_\tau + \sum_{j=1}^J \sum_{m=1}^M \mu_{j,m}\right)\tau\right\},$$

which implies that we can update τ by drawing

$$\tau \mid \cdot \sim \text{Gamma}\left(J\sum_{m=1}^{M}\nu_m + a_\tau - 1, \sum_{j=1}^{J}\sum_{m=1}^{M}\mu_{j,m} + b_\tau\right). \tag{3.22}$$

The full conditional of β is

$$\pi(\beta \mid \cdot) \propto \beta^{b_\beta - 1}\exp\left\{-\left[a_\beta + \sum_{i=1}^{M}\nu_m\right]\beta - E_1(\beta\nu_M)\right\}. \tag{3.23}$$

It is easy to show $\pi(\beta \mid \cdot)$ is a log concave function, thus, the ARS algorithm can be used to sample β.

3.2.2 HPGRF for multiple independent realizations

Although the HPGRF for a single realization discussed in Section 3.2.1 has a wide range of applications for multi-type spatial point processes, as pointed out in Section 3.1, a functional neuroimaging meta analysis involves more complicated multiple spatial point patterns, i.e. foci reported from different studies with different types. It is straightforward to assume that the foci from different studies are multiple independent realizations of a spatial point process for each type, where the spatial point processes are linked across types by shared clustering in certain regions of the brain. This motivates the need to extend the HPGRF for multiple independent realizations. Model properties (Theorem 1 – 3) do not change for each realization, but the posterior simulation algorithm need to be modified accordingly. We mainly focus on the extension based on the truncated model (3.13) as it can well approximate model (3.1).

Let $(\mathbf{x}_{i,j}, \mathbf{y}_{i,j})$, for $i = 1, 2, \ldots, n_j$, be multiple independent realizations of the Poisson point process $(\mathbf{X}_j, \mathbf{Y}_j)$ in model (3.13), where n_j is the number of realizations for $(\mathbf{X}_j, \mathbf{Y}_j)$. For each i and j, write $(\mathbf{x}_{i,j}, \mathbf{y}_{i,j}) = \{(\mathrm{x}_{i,j,l}, \mathrm{y}_{i,j,l})\}_{l=1}^{m_{i,j}}$, where $(\mathrm{x}_{i,j,l}, \mathrm{y}_{i,j,l})$ is an observed point in $(\mathbf{x}_{i,j}, \mathbf{y}_{i,j})$ indexed by l, for $l = 1, \ldots, m_{i,j}$, and $m_{i,j}$ is the observed number of points in $(\mathbf{x}_{i,j}, \mathbf{y}_{i,j})$. The joint density of $\{\{\mathbf{x}_{i,j}\}_{i=1}^{n_j}\}_{j=1}^{J}$, $\{\boldsymbol{\mu}_j\}_{j=1}^{J}$, $\{\sigma_j^2\}_{j=1}^{J}$, $\boldsymbol{\nu}$, $\boldsymbol{\theta}$, τ and β given $\{\{\mathbf{y}_{i,j}\}_{i=1}^{n_j}\}_{j=1}^{J}$ is proportional to

$$\begin{aligned}&\prod_{j=1}^{J}\left\{\prod_{i=1}^{n_j}\left[\pi(\mathbf{x}_{i,j}, \mathbf{y}_{i,j} \mid \boldsymbol{\mu}_j, \boldsymbol{\theta}, \sigma_j^2)\right]\pi(\sigma_j^2)\pi(\boldsymbol{\mu}_j \mid \boldsymbol{\nu}, \tau)\right\}\\ &\times\pi(\tau)\pi(\boldsymbol{\nu} \mid \beta)\pi(\beta)\pi(\boldsymbol{\theta}),\end{aligned} \tag{3.24}$$

where for each j the density $\pi(\mathbf{x}_{i,j}, \mathbf{y}_{i,j} \mid \boldsymbol{\mu}_j, \boldsymbol{\theta}, \sigma_j^2)$, for $i = 1, \ldots, n_j$ has the same form of $\pi(\mathbf{x}_j, \mathbf{y}_j \mid \boldsymbol{\mu}_j, \boldsymbol{\theta}, \sigma_j^2)$ in model (3.14), and all other parameters have the exact same density as the model (3.14). This implies that in the steps for sampling $\boldsymbol{\nu}$, τ and β are as described in 3.19, 3.22 and 3.23. For other parameters, we have the following sampling schemes.

- **Sampling $\mathbf{x}_{i,j}$**

  The full conditionals of $\mathrm{x}_{i,j,l}$, for $l = 1, \ldots, m_{i,j}$, are a similar to (3.15), i.e.

  $$\Pr(\mathrm{x}_{i,j,l} = \theta_m \mid \cdot) \propto \mu_{j,m} k_{\sigma_j^2}(\mathrm{y}_{i,j,l}, \theta_m). \tag{3.25}$$

- **Sampling $\boldsymbol{\theta}$**

  The full conditional density of $\boldsymbol{\theta}$ with respect to $\prod_{m=1}^{M} \tilde{\alpha}(d\mathrm{x})$ is

  $$\begin{aligned}&\pi(\boldsymbol{\theta} \mid \cdot)\\ &\propto \exp\left\{-\sum_{m=1}^{M}\sum_{j=1}^{J} K_{\sigma_j^2}(\mathcal{B}, \theta_m)\mu_{jm} n_j\right\}\prod_{j=1}^{J}\prod_{i=1}^{n_j}\prod_{l=1}^{m_{i,j}}\left[\sum_{m=1}^{M}\mu_{j,m} I_{\mathrm{x}_{i,j,l}}(\theta_m)\right].\end{aligned} \tag{3.26}$$

  Let $\tilde{\theta}_1, \ldots, \tilde{\theta}_{\tilde{M}}$ be $\tilde{M}$ distinct points in $\{\{\{\mathrm{x}_{i,j,l}\}_{l=1}^{m_{ij}}\}_{i=1}^{n_j}\}_{j=1}^{J}$. We first draw a random permutation of {1, …, M} denoted by $\{p_1, \ldots, p_M\}$. Then for $1 \leq m \leq \tilde{M}$, let $\theta_{p_m} = \tilde{\theta}_m$. For $m > \tilde{M}$, draw θ_{p_m} according to the following density with respect to $\tilde{\alpha}(d\mathrm{x})$,

  $$\pi(\theta_{p_m} \mid \cdot) \propto \exp\left\{-\sum_{j=1}^{J} K_{\sigma_j^2}(\mathcal{B}, \theta_{p_m})\mu_{j,p_m} n_j\right\}. \tag{3.27}$$

- **Sampling $\boldsymbol{\mu}_j$**

  The full conditional distribution of $\mu_{j,m}$, for $m = 1, \ldots, M$, is

  $$[\mu_{j,m} \mid \cdot] \sim \text{Gamma}\left[\nu_m + \sum_{i=1}^{n_j} \sum_{l=1}^{m_{i,j}} I_{\theta_m}(\mathrm{x}_{i,j,l}), n_j K_{\sigma_j^2}(\mathcal{B}, \theta_m) + \tau\right], \tag{3.28}$$

  which is a straightforward extension of Theorem 4.

- **Sampling σ_j^2**

  With the same assumptions on the prior as in Section 3.2.1, the full conditional density of σ_j^2 is, for $\sigma_j^2 \in [a_\sigma, b_\sigma]$,

  $$\begin{aligned} &\pi(\sigma_j^2 \mid \cdot) \\ &\propto \exp\left[-\sum_{m=1}^{M} \left\{K_{\sigma_j^2}(\mathcal{B}, \theta_m)\mu_{j,m} n_j\right\} - \frac{S_{\cdot,j}^2}{\sigma_j^2} - \frac{dm_{\cdot,j}}{2}\log(\sigma_j^2)\right], \end{aligned} \tag{3.29}$$

  where $S_{\cdot,j}^2 = \frac{1}{2}\sum_{i=1}^{n_j}\sum_{l=1}^{m_{i,j}} \|\mathrm{x}_{i,j,l} - \mathrm{y}_{i,j,l}\|^2$ and $m_{\cdot,j} = \sum_{j=1}^{n_j} m_{i,j}$. Thus, a similar random walk scheme can be used as in Section 3.2.1.

3.3 Examples

In this section, we illustrate the proposed HPGRF model on the bramble canes example and the neuroimaging meta-analysis example discussed in the Introduction. For both analyses, we assign vague priors on the hyper-parameters, i.e. $\sigma_j^{-2} \sim U[0, 10]$, $\beta \sim \text{Gamma}(0.001, 0.001)$ and $\tau \sim \text{Gamma}(0.001, 0.001)$. Also, the base measure $\alpha(d\mathrm{x}) = \Pi(d\mathrm{x})/|\mathcal{B}|$ where $\Pi(d\mathrm{x})$ denotes the Lebesgue measure and $|\mathcal{B}| = \int_{\mathcal{B}} \Pi(d\mathrm{x})$. This implies that $\tilde{\alpha}(d\mathrm{x}) = \alpha(d\mathrm{x})$ and $\tilde{\alpha}(\mathcal{B}) = 1$ (Theorem 2).

3.3.1 Bramble Canes

The bramble canes data were recorded and first analyzed by Hutchings (1979). The data set includes information on the locations and ages of bramble canes in a 9 meter squared field. For our analysis we estimate the intensity function on a

100×100 regular lattice covering the field. The canes are classified as either newly emergent or one year old, see Figure 3.1. The data were subsequently analyzed by Diggle (1981; 1983) and Diggle and Milne (1985). All these analyses indicate that each type of cane exhibits clustering. The multivariate J function proposed by van Lieshout and Baddeley (1999) can be used to measure the dependence between the two cohorts. In particular, we choose a cross type J function, i.e. J_{01}, see van Lieshout and Baddeley (1999) for the definition, which makes a comparison between the distributions of the distances to the nearest one-year old cane measured from any location in the region and a newly emergent cane. We consider the null hypothesis that the two types of bramble cane patterns are independent, in which case $J_{01} \equiv 1$. Figure 3.2 shows the Kaplan-Meier estimate of J_{01} for the bramble canes data and the 95% confidence envelope of the estimate under the null distribution, which are computed from 100 simulations obtained by wrapping the observation window on a torus and randomly translating the point pattern of newly emergent canes (Lotwick and Silverman 1982). In Figure 3.2, the estimated J_{01} is significantly smaller than 1 and falls outside the null confidence envelope. This implies that the presence of a newly emergent cane increases the chance of finding a one-year old cane nearby. However, to our best knowledge, no nonparametric methods exist that can estimate the intensity shared between types. We apply our HPGRF model to jointly analyze the two types of bramble canes, which can produce the intensity estimates for each type of bramble canes as well as a population mean intensity between the two types. Also, we apply the PGRF model to fit each type of point pattern separately, where the kernel function and the prior of the smoothing parameters are identical to those used in HPGRF, the base measure for each gamma random field is also set to be $\alpha(d\mathrm{x})$, and the scale parameters for both types are assumed to follow a $\mathrm{Gamma}(0.001, 0.001)$

distribution.

Posterior estimates are obtained from 50,000 iterations of simulation with a burn-in of 20,000 iterations. Figure 3.3 displays the estimated posterior mean intensity functions obtained from both models. Comparing the intensity estimates for both types of canes in Figure 3.3, the HPGRF intensity estimate is spatially more diffuse than the PGRF intensity estimate. Table 3.1 summarizes the posterior distribution of the smoothing kernel variance parameter σ_j^2 obtained from both models. The HPGRF estimate are much larger. As the multivariate J function suggests the dependence between the two cohorts, the HPGRF should provide more reliable results about the kernel parameters, since it borrows information across the two types, while the PGRF only estimates the kernel parameters using data within each type. In Figure 3, the intensity estimates for the two types of canes by HPGRF shows more similarity compared with the results estimated by the PGRF. Also, the HPGRF provides the population mean intensity estimates from which we can identify the common clustering regions between the two types of canes.

3.3.2 Functional Neuroimaging Meta Analysis of Emotion

In this section, we apply the proposed model to a functional neuroimaging meta analysis of emotion first reported by Kober et al. (2008) and further analyzed in Chapter II. This data set consists of 164 publications of various aspects of emotion. Kober et al. (2008) proposed a multi-level kernel density analysis (MKDA) approach. The goal is to identify consistent regions of activation across the different studies and types of emotion. In Chapter II, we address this problem using a Bayesian hierarchical independent cluster process (BHICP) model. However, both MKDA and BHICP only analyze a single group of spatial point patterns that ignores the different types of emotions studied and focus on finding consistent activation over the different types of emotions. It is also of great interest to identify consistent activation regions between sub types, such as the type of emotion. The HPGRF model can take into correlation between sub types and identify consistent activations regions that are shared by different sub types. In particular, we focus on a subset of the meta-analysis dataset, including 219 studies reporting 1393 foci of five emotions: sad, happy, anger, fear and disgust, as shown in Figure 3.4. Table 3.2 lists some features and summary statistics of this data set. Posterior estimates are based on 120,000 iterations of simulation with a burn-in of 20,000, The posterior mean intensity estimate are presented in Figure 3.5. We also separately fit a spatial point process for each sub population, although the sub populations might be correlated. However, the original PGRF (Wolpert and Ickstadt 1998a) cannot be directly applied to our data, as it is only designed for a single realization of a spatial point process. Thus, we modify the PGRF model to fit multiple independent realizations of a spatial point process. We refer to this model as IPGRF. See Appendix B for more details. We compare the intensity estimates between the HPGRF and the IPGRF for two axial slices $\mathrm{Z} = -20, 26\mathrm{mm}$, as shown in Figure 3.6, from which we can see the intensity estimates by HPGRF and IPGRF are qualitatively similarly. The HPGRF intensity estimate is a bit more spatially diffuse than the IPGRF intensity estimate. Table 3.4 shows the smoothing kernel variances estimated by the two methods, where the HPGRF estimates are slightly larger than IPGRF estimates but the 95% credible intervals between the two estimates overlap. As shown in Figure 3.5, the HPGRF can estimate the "population level" intensity which finds consistent activation regions across the different emotions. We can see there is a high and concentrated intensity in axial slice $Z = 15\mathrm{mm}$ in the "population level" intensity. It is actually in the right amygdala. The amygdala, an almond-sized and -shaped brain structure, has been shown to be linked with a person's mental and emotional state (Adolphs 1999). Our analysis indicates that the five emotions share high activation in the right amygdala.

3.4 Simulation Studies

In this section, we report results from simulation studies that demonstrate the ability of the HPGRF to borrow information from different types and provide more accurate intensity estimates compared to the IPGRF model.

3.4.1 Simulated Data Sets

We simulate 2D spatial point patterns on region $A = [0, 100]^2$ from three different types of points based on Poisson point processes with mixture intensity functions. i.e., for $i = 1, \ldots, N$,

$$\mathbf{Y}_{i,j} \mid \boldsymbol{\mu}, \boldsymbol{\Sigma} \sim \mathcal{PP}\{A, \lambda_j(\mathrm{x})d\mathrm{x}\}, \quad j = 1, 2, 3. \tag{3.30}$$

where $\lambda_j(\mathrm{x}) = \epsilon + \sum_{(\theta,\mu,\Sigma)\in(\boldsymbol{\theta},\boldsymbol{\mu},\boldsymbol{\Sigma})_j} \theta\phi_2(\mathrm{x};\mu,\Sigma)$ and $\phi_d(\mathrm{x};\mu,\Sigma)$ denotes the d-dimensional Gaussian density at x with mean μ and covariance Σ. We set intensity parameters such that point patterns from different types show clustering on four regions, where the intensity parameters are summarized in Table 3.3 and where $\epsilon = 0.001$, We assume that the three sub types have intensity functions:

$$\begin{array}{rcllll}
\lambda_1(\mathrm{x}) & = & \epsilon & + \theta_2\phi_2(\mathrm{x};\mu_2,\Sigma_2) & + \theta_3\phi_2(\mathrm{x};\mu_3,\Sigma_3) & , \\
\lambda_2(\mathrm{x}) & = & \epsilon & + \theta_2\phi_2(\mathrm{x};\mu_2,\Sigma_2) & & + \theta_4\phi_2(\mathrm{x};\mu_4,\Sigma_4), \\
\lambda_3(\mathrm{x}) & = & \epsilon + \theta_1\phi_2(\mathrm{x};\mu_1,\Sigma_1) & + \theta_2\phi_2(\mathrm{x};\mu_2,\Sigma_2) & + \theta_3\phi_2(\mathrm{x};\mu_3,\Sigma_3) & .
\end{array}$$

The true intensity functions are presented in Figure 3.7. This shows that the three sub types share clustering in region 2 and type 1 and type 3 share clustering in region 3. Also, only points from type 1 and type 3 cluster about region 1 and region 4, respectively.

3.4.2 Accuracy of Intensity Estimates

We simulate $K = 1000$ data sets according to the model specifications in the previous section and fit each data sets via the HPGRF and the IPGRF models respectively. Figure 3.7 present the estimated posterior mean intensity functions for one of the simulated data sets compared with the true intensity. We measure the model performance based on the sub-type average integrated mean square error (IMSE) and integrated weighted mean square error (IWMSE), i.e.

$$\begin{aligned} IMSE &= \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}\int_{A}[\tilde{\lambda}_{jk}(\mathrm{x}) - \lambda_j(\mathrm{x})]^2 d\mathrm{x} \\ IWMSE &= \frac{1}{JK}\sum_{j=1}^{J}\sum_{k=1}^{K}\int_{A}[\tilde{\lambda}_{jk}(\mathrm{x}) - \lambda_j(\mathrm{x})]^2 \lambda_j(\mathrm{x}) d\mathrm{x} \end{aligned}$$

where $\tilde{\lambda}_{jk}(\mathrm{x})$ is the estimated posterior mean intensity function for type j in simulation k and $\lambda_j(\mathrm{x})$ is the true intensity function. IWMSE gives more weight to regions with a large true intensity. The results are summarized in Table 3.5 which shows that the HPGRF results in small IMSE and IWMSE that are 27% and 35% smaller than those of the IPGRF, respectively. Thus, when sub types share clustering. The HPGRF model is more accurate.

3.5 Discussion

In this article, we focus on the analysis of multi type spatial point patterns to which we propose a nonparametric Bayesian approach, i.e. the HPGRF, by extending the PGRF model by Wolpert and Ickstadt (1998a). The HPGRF model accounts for the dependence in the point patterns between the different types. The HPGRF is able to deal with more complicated multivariate spatial point patterns than the standard PGRF can handle, i.e. multiple groups of spatial point patterns, where each group contains one or more independent realizations of a spatial point process. The spatial point processes across groups can either be independent of one another or remain connected, depending on the particular features of the data. The real data analyses suggest that our model can provide "population mean" intensity estimates to identify common regions that share clustering providing better interpretation of data. The simulation studies shows that the HPGRF can improve the accuracy of intensity estimate for each sub type.

There are several future directions. First, the HPGRF can be further extended to more than two levels of hierarchy for more complex spatial point patterns, where the depth depends on the needs of the data analysis. Another interesting extension is to make the HPGRF model able to deal with multiple dependent realizations of multi-type spatial point processes, which is potentially useful for spatial-temporal data. Also, the current posterior computational method is based on a truncation method, where the number of gamma points, i.e. M, must be much larger than the observed number of points. This motivates the needs of more efficient posterior simulation algorithm for massive data sets in the future.

Table 3.1: Summary statistics of the posterior distribution of the smoothing kernel variance σ_j^2 via HPGRF and PGRF for bramble canes data.

Bramble canes	Statistics	HPGRF	PGRF
Newly emergent	Mean.	23.7	3.1
	Sd.	4.1	0.6
	C.I.	(16.9,32.4)	(2.2, 4.5)
One year old	Mean.	28.6	6.9
	Sd.	5.6	1.4
	C.I.	(20.1,42.3)	(4.6, 9.9)

Table 3.2: Number of studies and foci in the emotion emotion analysis data set.

	sad	happy	anger	fear	disgust
Studies	45	36	26	68	44
Foci	346	177	166	367	337

Table 3.3: Parameters for true intensity functions

Region j	1	2	3	4
σ_j	15	10	5	10
μ_j	$(60,75)^T$	$(70,30)^T$	$(40,50)^T$	$(10,20)^T$
Σ_j	$\begin{pmatrix} 30 & 15 \\ 15 & 15 \end{pmatrix}$	$\begin{pmatrix} 30 & -10 \\ -10 & 40 \end{pmatrix}$	$\begin{pmatrix} 20 & -5 \\ -5 & 10 \end{pmatrix}$	$\begin{pmatrix} 10 & 5 \\ 5 & 20 \end{pmatrix}$

Table 3.4: Summary statistics of the posterior distribution of the smoothing kernel variance σ_j^2 via HPGRF and PGRF for five emotion meta analysis

Methods	Statistics	Emotions				
		Sad	Happy	Anger	Fear	Disgust
HPGRF	Mean.	0.94	1.02	1.07	1.01	1.08
	C.I.	(0.87,1.02)	(0.90,1.14)	(0.94,1.19)	(0.93,1.09)	(1.00,1.18)
IPGRF	Mean	0.97	0.99	0.99	0.97	0.98
	C.I.	(0.94,0.99)	(0.97,1.01)	(0.97,1.01)	(0.95,0.99)	(0.96,1.00)

Table 3.5: The IMSE and IWMSE of the intensity estimation via HPGRF and IPGRF

$\times 10^{-3}$	HPGRF	IPGRF
IMSE (s.e.)	176.27(22.84)	227.42(25.74)
IWMSE (s.e.)	11.17(2.64)	17.09(2.62)

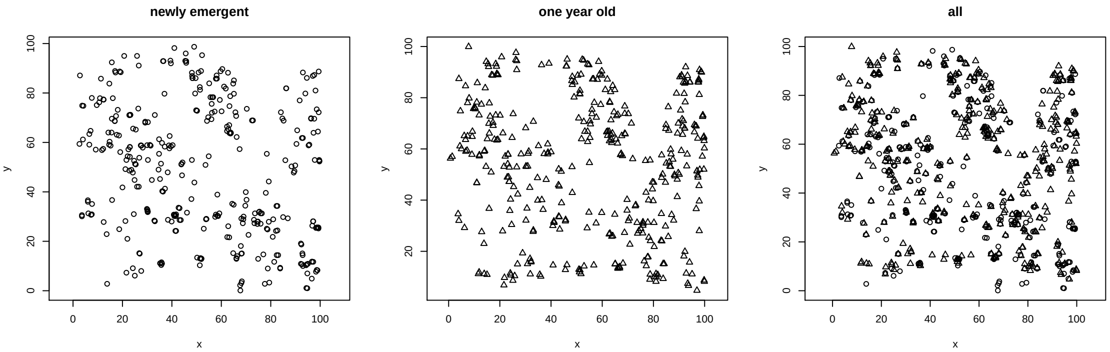


Figure 3.1: Locations of newly emergent and one year old bramble canes.

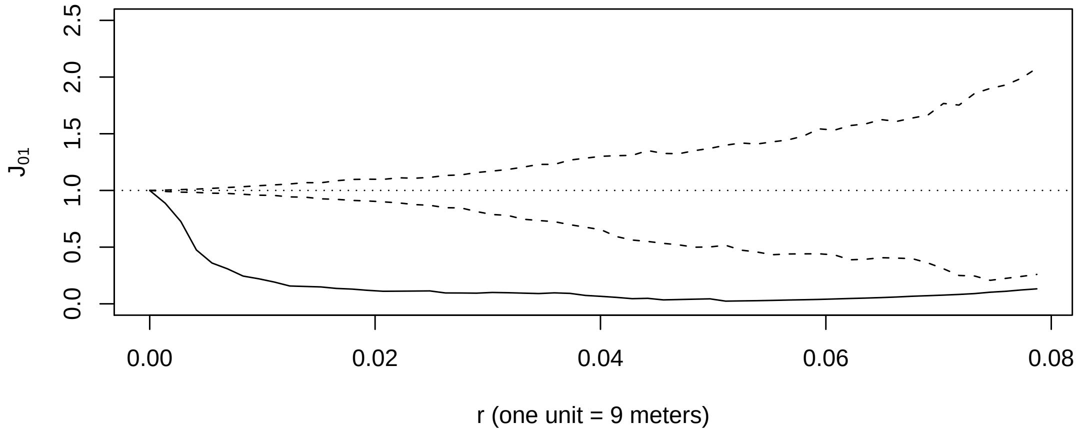


Figure 3.2: The estimate of the multivariate J function, e.g. $J_{01}(r)$ (solid line) for the bramble canes data with the 95% confidence envelope (dashed lines) based on 100 random torus translations of the data.

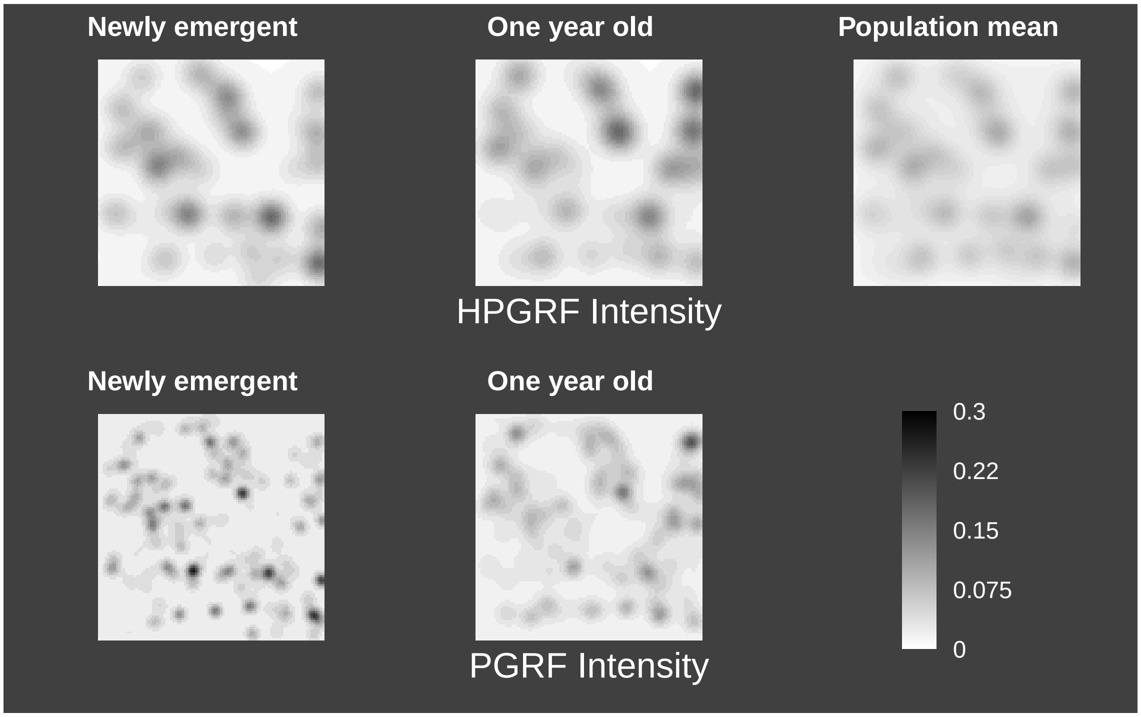


Figure 3.3: The HPGRF and PGRF posterior mean intensity estimates for each type of bramble canes point patterns. The HPGRF also provide a "population mean" intensity estimate.

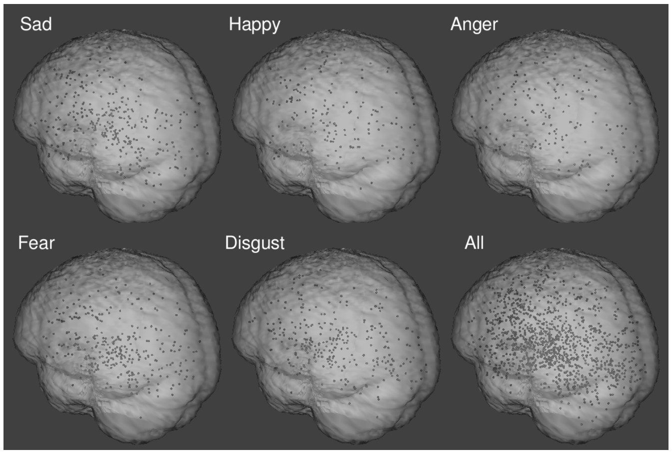


Figure 3.4: The 1393 foci reported from 219 studies of five emotions

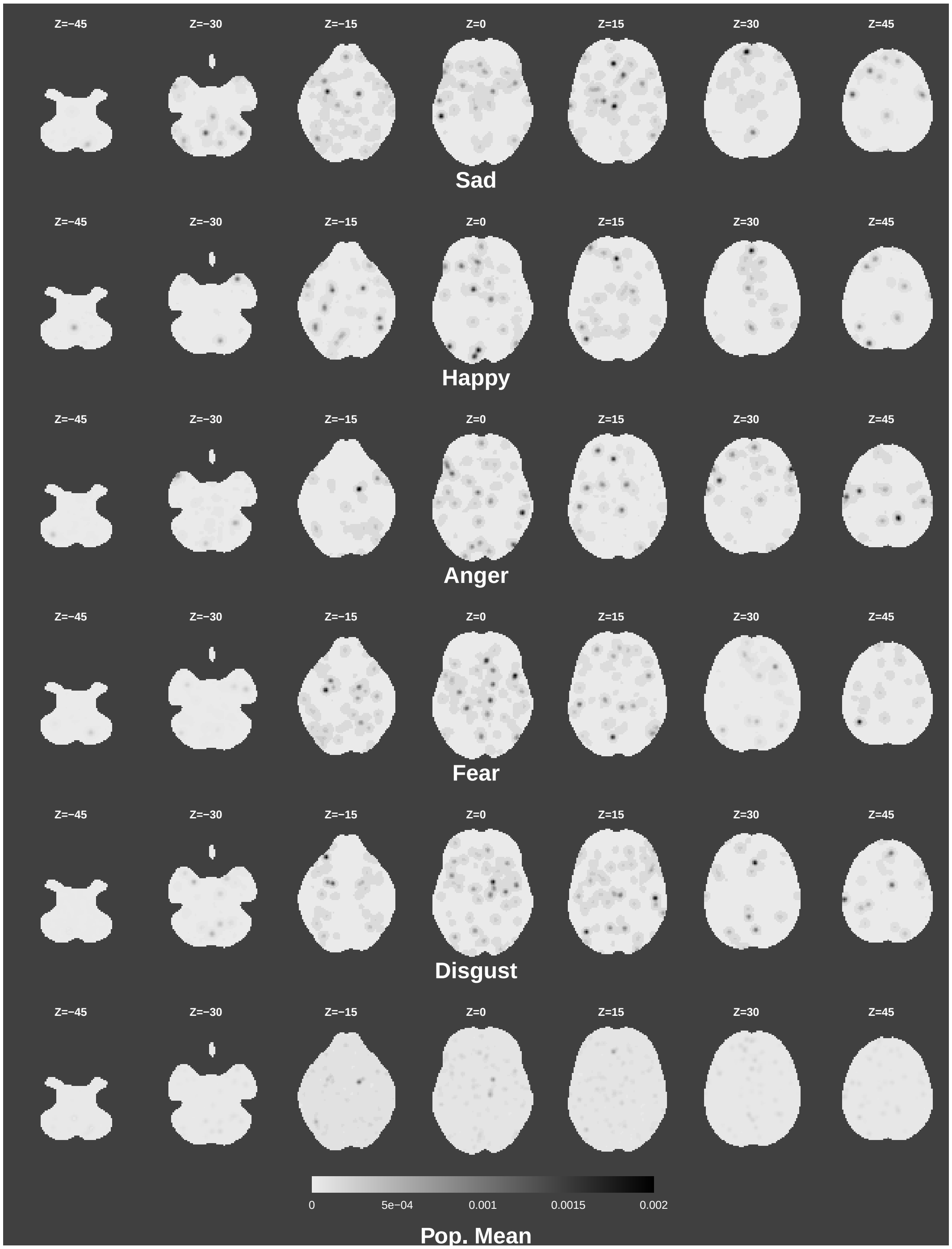
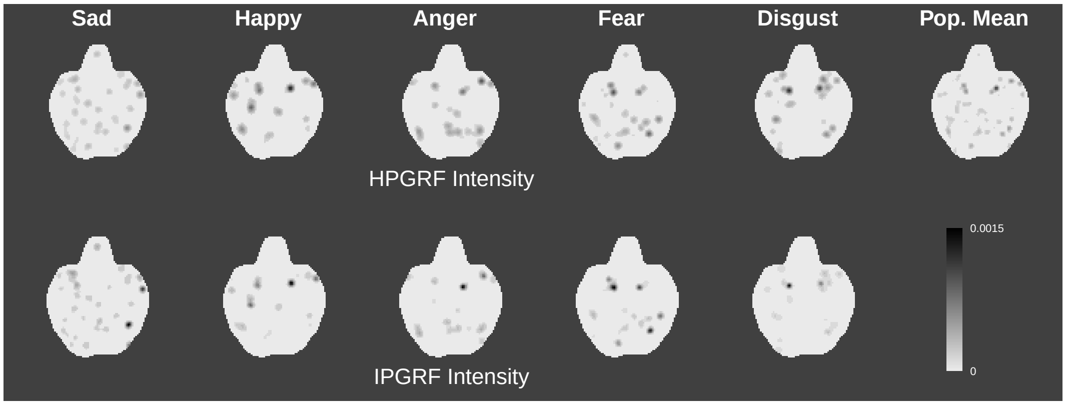


Figure 3.5: HPGRF: Foci intensity estimation for five emotions (Sad, Happy, Anger, Fear and Disgust) compared with the "population mean" intensity estimation

$Z = -20$mm



$Z = 26$mm

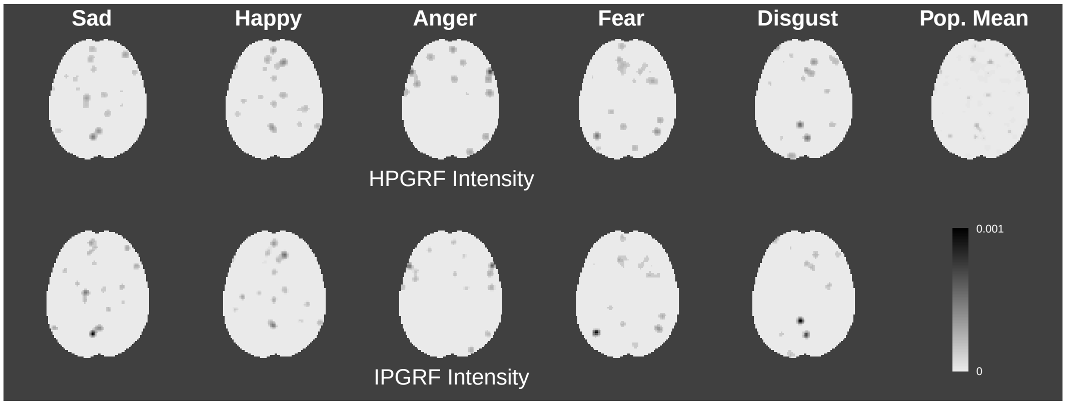


Figure 3.6: The HPGRF posterior mean intensity estimates of the two axial slices ($Z = -20$mm, 26mm) for five emotions (Sad, Happy, Anger, Fear and Disgust) compared with the IPGRF intensity estimates.

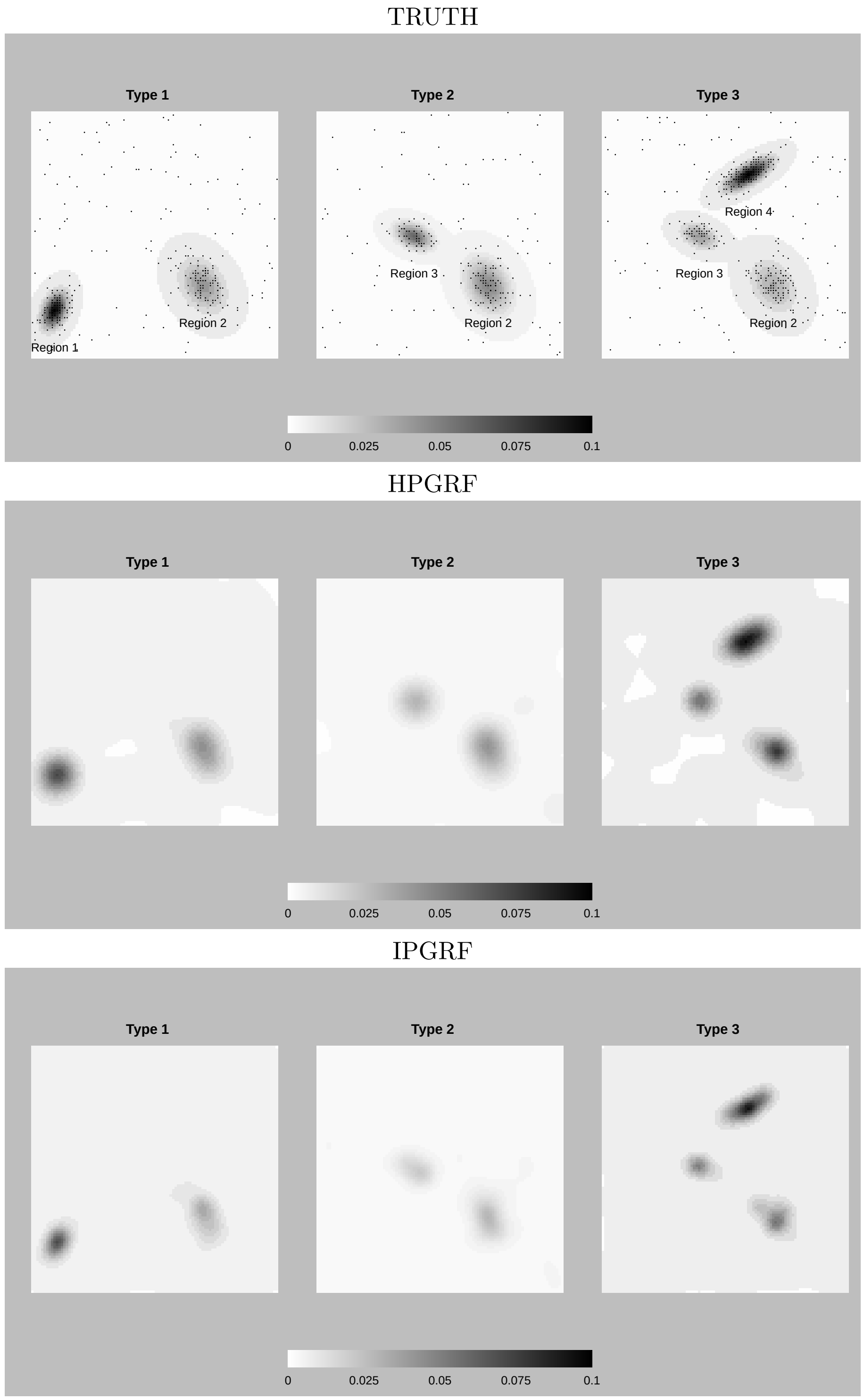


Figure 3.7: The true intensity compared with the HPGRF and IPGRF intensity estimates for the simulated Poisson cluster point process data sets.

CHAPTER IV

A Bayesian Spatial Point Process Classifier with Application to Functional Neuroimaging Reverse Inference

Decoding of behavioral information or cognitive states from human brain activity is a primary goal in functional neuroimaging research. Many studies are designed to determine the distributed patterns of brain activation that result from different known brain functions. In the meta analysis of these studies, it is of great interest to perform reverse inference on cognitive states from peak activation locations that are reported from a new study. Motivated by this problem, we develop a Bayesian classification model for multiple types of point patterns via spatial point processes. With appropriate modeling of the intensity function for each type of point pattern, we construct a classifier based on the posterior probability of the type given a point pattern. We show this posterior probability, for a new study, can be efficiently estimated via importance sampling. We also propose a statistical map as an exploratory tool to discover the impact of brain regions on distinguishing cognitive states. Our method is illustrated by the analysis of 437 emotion studies. Our model shows a higher correct classification rate compared with a naive Bayes classifier, and the proposed statistical map provides valuable insights into the relationships between emotions and brain activity.

4.1 Introduction

With the rapid growth of functional neuroimaging studies during the past two decades researchers have shown that there are strong connections between mental states and brain activity. Most functional neuroimaging studies are designed to address this problem based on a "forward inference" procedure. Identify the brain activity that results from given psychological tasks. A key challenge for cognitive neuroscience is to decode metal states based on the measurements of brain activity that are obtained from functional neuroimaging techniques. We refer to this procedure as "reverse inference". There is an increasing interest in developing statistical machine learning approaches to reverse inference using functional Magnetic Resonance Imaging (fMRI) data, see Pereira et al. (2009) for a tutorial overview. Most of these methods are developed particularly for modeling subject level statistical parametric maps (SPM) or voxel-wise time series that are acquired from a single fMRI study given known tasks or metal states. However, due to small sample size, single fMRI studies usually suffer from low power and poor reproducibility (Yarkoni 2009, Kober et al. 2008, Eickhoff et al. 2009). To overcome these limitations, researchers are paying much attention to formal synthesis of cognitive neuroscience literature via statistical meta analysis (Yarkoni et al. 2010, Kang et al. 2011). Current methods for meta-analysis, like Activation likelihood Estimation (ALE) by Eickhoff et al. (2009), Multi-Kernel Density Analysis (MKDA) by Kober et al. (2008), Bayesian Hierarchical Independent Cluster Process (BHICP) in Chapter II and Hierarchical Poisson/Gamma Random Fields (HPGRF) in Chapter III, primarily focus on forward inference, i.e. localizing consistent effects over studies using reported peak activation foci. The MKDA and the ALE are mass-univariate approaches and require fixed tuning parameters, Furthermore, they also do not provide an interpretable fitted model, and cannot produce spatial confidence intervals on locations of activation. The BHICP and the HPGRF are fully Bayesian spatial models, where BHICP explicitly models the clustering of study foci about latent activation centers (accounting for multiple peaks within a study describing a single anatomical region), which in turn cluster about latent population centers. In particular, the BHICP can dissociate between inter-study spread of foci and spatial uncertainty in population response location. The HPGRF is a non-parametric Bayesian approach that jointly analyses multiple types of foci, where each group is associated with a particular task or psychological state. It assumes that the random intensities for different groups are linked, which allows sharing of clusters between different types.

Although meta analysis offers a unique opportunity to make psychological inferences across many task domains, there is little work aimed at reverse inference. Yarkoni et al. (2011) combine a naive Bayes classifier (NBC) and the MKDA. For each study, this method creates binary activation maps using the MKDA, with a value of 1 (activated) assigned to each voxel in the brain if it is within a certain distance (a spherical kernel size) of a reported focus, and 0 (non-activated) otherwise. These binary activation maps are in turn treated as feature variables in the naive Bayes classifier. The study type, i.e. the designed psychological state, determines the class membership. More specifically, for each type, an activation probability map is constructed by a taking weighted average of the binary maps. Using the activation probability maps, the predictive probability of the study type given activation from a new study is then computed based on Bayes' theorem under the independence assumption (see Yarkoni et al. (2011) for details). This method is very computational efficient and can handle extremely large sets of voxels without difficulty. Also, it performs well in classifying working memory from emotion from pain studies. However, there are several potential drawbacks of this method. First, it ignores the spatial dependence in the activation maps, leading to biased predictive probabilities of the class membership. Second, it requires fixed tuning parameters that might affect the classification performance: the current method simply sets these parameters based on experience rather than estimating them from data. Third, it only focuses on the difference in the spatial distributions of foci between groups while neglecting the absolute rates of foci, which may be important for classification.

To address the above issues, we propose a Bayesian spatial point process based classifier that provides a general modeling framework for the classification of multiple types of spatial point patterns. More specifically, for each study, we observe a set of foci that are assumed to be a single realization of a spatial point process. Also, we have information about the type of each study. Thus, the foci reported from different studies can be modeled as multiple independent realizations of a multi-type point process. The classifier is, in turn, constructed from the posterior predictive probability of the type of a new study. Although our method can be applied to any model that has an explicit intensity or density function, its application may be non-trivial for some complex hierarchical models.

In this article, we discuss how to apply our method to build up a Bayesian classifier based on the BHICP or the HPGRF. Our model has at least three advantages compared with the MKDA based NBC. First, in meta analysis data, the number of foci reported, as well as their locations are random. Both the random number and random locations are explicitly modeled by in our spatial point process models. Furthermore, spatial dependence between foci is also modeled. These features of the data are not modeled by the MKDA based NBC. Second, our hierarchical spatial model is a more accurate representation of the true data generating process, relative to the MKDA kernel mapping of points to voxel-wise images. Third, our fully Bayesian model captures more sources of variation and appropriately conveys the certainty in the computation of the predictive probabilities that determines the classification outcome.

The performance of our method is evaluated via leave-one-out cross validation (LOOCV) on all the observed data. Specifically, leaving one study out, we conduct Bayesian learning of the parameters on the foci reported from the remaining studies and make a prediction on the study type for the study left out. We repeat the procedure for each study and compute the classification rate based on all the observed studies. The above procedure for a Bayesian model can be very computationally expensive because it involves multiple posterior simulations. We employ an importance sampling approach to reduce the computation of LOOCV for our models. This idea has been investigated by many researchers (Gelfand et al. 1992, Gelfand 1996, Alqallaf and Gustafson 2001, Vehtari and Lampinen 2002) for a general Bayesian model. We apply this idea to the proposed Bayesian spatial point process classifier to avoid multiple posterior simulations. Due to the complexity of the model, this is not a trivial application of important sampling, Nevertheless, the gain in computational efficiency is great.

For decoding of cognitive states from brain activity, neuropsychologists are also interested in understanding the role of individual locations of the brain in the classification. To address this problem, the MKDA based NBC method can provide a voxel-wise reverse inference probability map, called a salience map, where the value on each map represents the probability of the study type given a focus is reported at that voxel. Since our model simultaneously considers the entire point pattern from a study, we see no unique way to measure the contribution of a single voxel. Instead, we propose a single foci impact map (SFIM), as an exploratory tool to discover the impact of brain regions in the classification. Given a particular study, for each type, the value of study-level SFIM at each voxel represents the change in the probability of the group by adding that voxel (or a focus located in that voxel) into the observed point pattern from that study. The group-level SFIM for each type is defined as an average of the study-level SFIMs over all possible studies that can be generated from the fitted model. As illustrated in Section 4.4, SFIM provides valuable insight into the relationships between psychological states and brain activity.

The reminder of this chapter is organized as follows. In Section 4.2, we first describe a general Bayesian model for reverse inference in neuroimaging meta analysis, then we incorporate the BHICP and the HPGRF into this modeling framework. Also, we discuss some computational details of LOOCV. In Section 4.3, we propose the single focus impact map to understand an individual voxel's contribution to classification. We demonstrate our methods in Section 4.4 on a meta-analysis of emotions, classifying different sub-types of emotion, and compare to the MKDA based NBC. In the last section, we conclude our work with a discussion and ideas for future research.

4.2 The Model

Suppose we have observed data $\mathcal{D}_n = \{(\mathbf{x}_i, t_i)\}_{i=1}^n$ consisting of n studies where $\mathbf{x}_i$ denotes the observed point pattern in the brain reported from study i, $t_i \in \{1, \ldots, J\}$ is the study type and J is the number of study types. Our goal is to make inference on the study type for a new study, i.e., t_{n+1} given the foci $\mathbf{x}_{n+1}$ obtained from this study and the observed data $\mathcal{D}_n$. Next, we describe a general Bayesian modeling framework to build a spatial point process classifier.

4.2.1 Bayesian Spatial Point Process Classifier

For each $i = 1, 2, \ldots, n+1$, we assume that $\mathbf{x}_i$ is a single realization of a spatial point process $\mathbf{X}_i$ and t_i is a sample realization of the type indicator, T_i. We infer on T_{n+1} based on the following posterior predictive probability, for $j = 1, \ldots, J$,

$$\begin{aligned} &\Pr(T_{n+1} = j \mid \mathbf{x}_{n+1}, \mathcal{D}_n) = \\ &\int \Pr(T_{n+1} = j \mid \Theta, \mathbf{x}_{n+1}, \mathcal{D}_n)\pi(\Theta \mid \mathbf{x}_{n+1}, \mathcal{D}_n) d\Theta, \end{aligned} \tag{4.1}$$

where Θ includes all the parameters involved in the sampling model of data $\pi(\mathcal{D}_{n+1} \mid \Theta)$. The estimate of T_{n+1} is usually given by

$$\hat{T}_{n+1} = \arg\max_j \Pr(T_{n+1} = j \mid \mathbf{x}_{n+1}, \mathcal{D}_n). \tag{4.2}$$

To obtain (4.1), we need to have the specification of $\Pr(T_{n+1} = j \mid \Theta, \mathbf{x}_{n+1}, \mathcal{D}_n)$ and $\pi(\Theta \mid \mathbf{x}_{n+1}, \mathcal{D}_n)$, which can be derived from the sampling model $\pi(\mathcal{D}_{n+1} \mid \Theta)$ and the prior distribution $\pi(\Theta)$ based on Bayes' theorem. A widely used assumption on $\pi(\mathcal{D}_{n+1} \mid \Theta)$ is the conditional independence of different samples given the parameters (Gelman et al. 2004), i.e.

$$\pi(\mathcal{D}_{n+1} \mid \Theta) = \prod_{i=1}^{n+1} \pi(\mathbf{x}_i, t_i \mid \Theta). \tag{4.3}$$

This implies that $\Pr(T_{n+1} = j \mid \Theta, \mathbf{x}_{n+1}, \mathcal{D}_n) = \Pr(T_{n+1} = j \mid \Theta, \mathbf{x}_{n+1})$. Therefore, the most important problem is to specify the joint distribution of $\mathbf{X}_i$ and T_i given Θ. In general, there are two approaches to modeling $\pi(\mathbf{x}_i, t_i \mid \Theta)$, which can be either written as $\Pr(T_i = t_i \mid \Theta)\pi(\mathbf{x}_i \mid T_i = t_i, \Theta)$ or $\pi(\mathbf{x}_i \mid \Theta)\Pr(T_i = t_i \mid \mathbf{x}_i, \Theta)$. In the former approach, we first specify the distribution of the study type and then model the distribution of foci given the study type, while in the latter approach we model foci before the type. Although both approaches are commonly used in general classification problems (Hastie et al. 2008), we feel the former approach provides a more accurate representation of the true data generating process in our problem since neuroimaging studies selected in the meta analysis are all designed to observe the brain activity pattern given the pre-specified task or the psychological states. Thus, we need to specify $\Pr(T_i = t_i \mid \Theta)$, the a priori probability of the study type, as well as $\pi(\mathbf{x}_i \mid T_i = t_i, \Theta)$: the sampling distribution of the foci from each study given their study type. In summary, we consider the following general model, for $i = 1, \ldots, n+1$,

$$\begin{aligned} &\mathbf{X}_i \mid T_i = j, \Theta \overset{i.i.d.}{\sim} \text{Poisson}[\mathcal{B}, \lambda_j(x \mid \Theta)] \\ &\Pr[T_i = j] = p_j, \qquad \text{for } j = 1, \ldots, J, \\ &\Theta \sim \pi(\Theta), \end{aligned}$$

where $\text{Poisson}[\mathcal{B}, \lambda_j(x \mid \Theta)]$ denotes a Poisson point process on $\mathcal{B}$ with intensity function $\lambda_j(x \mid \Theta)$. The parameter p_j is pre-specified and represents the prior probability for study type j. Θ includes all other parameters or latent quantities that are involved in the model. The prior density of Θ is denoted $\pi(\Theta)$. To estimate the posterior predictive distribution (4.1) for T_{n+1}, one may directly draw Θ from $\pi(\Theta \mid \mathbf{x}_{n+1}, \mathcal{D}_n)$. However, this is not straightforward to implement. Thus, we consider an alternative approach using the posterior draw of Θ given $\mathcal{D}_n$ and then use an importance sampling approach. We have the follow proposition:

Proposition 1. *The posterior predictive distribution for T_{n+1} is given by*

$$\begin{aligned} &\Pr[T_{n+1} = j \mid \mathbf{x}_{n+1}, \mathcal{D}_n] = \\ &\qquad \frac{p_j \int \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j, \Theta) \pi(\Theta \mid \mathcal{D}_n) d\Theta}{\sum_{j'=1}^{J} p_{j'} \int \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j', \Theta) \pi(\Theta \mid \mathcal{D}_n) d\Theta} \end{aligned} \tag{4.4}$$

where

$$\pi(\mathbf{x}_{n+1} \mid T_{n+1} = j, \Theta) \;=\; \exp\left\{|\mathcal{B}| - \int_{\mathcal{B}} \lambda_j(x \mid \Theta) dx\right\} \prod_{x \in \mathbf{x}_{n+1}} \lambda_j(x \mid \Theta).$$

This proposition leads to the following algorithm used to estimate the posterior predictive probability used for reverse inference.

Reverse inference algorithm

- **Input**: The observed data $\mathcal{D}_n$, a foci pattern, $\mathbf{x}_{n+1}$, reported from a new study and the total number of simulations K.
- **Step 1**: Run a Bayesian spatial point process model to obtain the posterior draws $\Theta^{(k)} \sim \pi(\Theta \mid \mathcal{D}_n)$, for $k = 1, \ldots, K$.
- **Step 2**: Compute

  $$\pi_j^{(k)} = \exp\left\{-\int_{\mathcal{B}} \lambda_j(x \mid \Theta^{(k)}) dx\right\} \left\{\prod_{x \in \mathbf{x}_{n+1}} \lambda_j(x \mid \Theta^{(k)})\right\}.$$

- **Output**: The posterior predictive probability is given by

  $$\widehat{\Pr}[T_{n+1} = j \mid \mathbf{x}_{n+1}, \mathcal{D}_n] = \frac{p_j \sum_{k=1}^K \pi_j^{(k)}}{\sum_{j'=1}^J p_{j'} \sum_{k=1}^K \pi_{j'}^{(k)}}. \tag{4.5}$$

  The prediction of T_{n+1} is given by

  $$\hat{T}_{n+1} = \arg\max_j \left(p_j \sum_{k=1}^K \pi_j^{(k)}\right). \tag{4.6}$$

To evaluate the performance of our method, we are interested in computing the leave-one-out cross-validation (LOOCV) classification rates. More specifically, we use data $\mathcal{D}_{-i} = \{(\mathbf{x}_l, t_l)\}_{l \neq i}$ to make a predication of T_i denoted by $\hat{T}_i$, and focus on the $J \times J$ LOOCV confusion matrix $\boldsymbol{C} = \{c_{jj'}\}$, defined by

$$c_{jj'} = \frac{\sum_{i=1}^n I_j(t_i) I_{j'}(\hat{T}_i)}{\sum_{i=1}^n I_j(t_i)}, \tag{4.7}$$

where $I_a(b)$ is an indicator function. $I_a(b) = 1$ if $a = b$, $I_a(b) = 0$, otherwise. Then the overall and the average correct classification rates are respectively given by

$$c_o = \frac{1}{n}\sum_{i=1}^{n} I_{t_i}(\hat{T}_i), \quad \text{and} \quad c_a = \frac{1}{n}\sum_{i=1}^{n} c_{jj}. \tag{4.8}$$

In order to obtain $\hat{T}_i$, we compute the LOOCV predictive probabilities. For $i = 1, \ldots, n$,

$$\Pr[T_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}] = \int \Pr[T_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}, \Theta]\pi(\Theta \mid \mathbf{x}_i, \mathcal{D}_{-i})d\Theta, \tag{4.9}$$

which can be estimated via Monte Carlo simulation. However, it is not straightforward and very inefficient to draw Θ from $\pi(\Theta \mid \mathbf{x}_i, \mathcal{D}_{-i})$ for each i. Thus, to avoid having to run the multiple posterior simulations, we consider another representation of (4.9) in the following proposition.

Proposition 2. *The LOOCV predictive probabilities of T_i, for $i = 1, \ldots, n$, is given by*

$$\Pr[T_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}] = \frac{p_j Q_{jt_i}}{\sum_{j'=1}^{J} p_{j'} Q_{j't_i}}, \tag{4.10}$$

where

$$Q_{jj'} = \int \frac{\pi(\mathbf{x}_i \mid T_i = j, \Theta)}{\pi(\mathbf{x}_i \mid T_i = j', \Theta)} \pi(\Theta \mid \mathcal{D}_n)d\Theta.$$

Proposition 2 leads to the following algorithm.

LOOCV algorithm

- **Input**: The observed data $\mathcal{D}_n$ and the total number of simulations K.
- **Step 1**: Run a Bayesian spatial point process model to obtain the posterior draws $\Theta^{(k)} \sim \pi(\Theta \mid \mathcal{D}_n)$, for $k = 1, \ldots, K$.

- **Step 2**: For $i = 1, \ldots, n$ and $j = 1, \ldots, J$, compute

$$\widehat{Q}_{jt_i} = \frac{1}{K} \sum_{k=1}^{K} \frac{\pi(\mathbf{x}_i \mid T_i = j, \Theta_i^{(k)})}{\pi(\mathbf{x}_i \mid T_i = t_i, \Theta_i^{(k)})}. \tag{4.11}$$

- **Output**: The posterior of the predictive probabilities of T_i, for $i = 1, \ldots, n$, are given by

$$\widehat{\Pr}[T_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}] = \frac{p_j \hat{Q}_{jt_i}}{\sum_{j'=1}^{J} p_{j'} \hat{Q}_{j't_i}}. \tag{4.12}$$

  And the estimate of T_i is

$$\hat{T}_i = \arg\max_j \left(p_j Q_{jt_i} \right). \tag{4.13}$$

Both BHICP and HPGRF provide an explicit form of $\pi(\mathbf{x}_i \mid T_i = j, \Theta)$ for $i = 1, \ldots, n+1$ and the posterior estimates for $\{\Theta^{(k)}\}_{k=1}^{K}$. Hence, we now discuss details on how to build up Bayesian classifiers and the corresponding LOOCV algorithms from these two models and compare the results.

4.2.2 The BHICP Model

In this section, with the assumption of independence between different types, we provide a representation of the BHICP model for a meta analysis involving multi type foci. We assume that $\mathbf{x}_i$, the observed foci from study i, is one realization of an independent cluster process $\mathbf{X}_i$ on the brain, $\mathcal{B}$, with intensity λ_{1j}. Let $\mathbf{Y}_{ij}$ be an independent cluster process with intensity function λ_{2j} representing the latent study activation centers from study i of type j, and $\mathbf{Z}_j$ be the population centers of type j with constant intensity β_j. Define $\mathbf{\Psi}_{ij} = \{\Psi_y, y \in \mathbf{Y}_{ij}\}$ and $\mathbf{\Sigma}_j = \{\Sigma_z, z \in \mathbf{Z}_j\}$, which are the covariance matrices attached to $\mathbf{Y}_{ij}$ and $\mathbf{Z}_j$ to describe the cluster shapes. Write $\Theta_{ij} = \{\epsilon_{1ij}, \epsilon_{2ij}, \theta_{1ij}, \theta_{2ij}, \eta_{ij}, \mathbf{y}_{ij}, \mathbf{\Psi}_{ij}\}$ that collects all the unknown parameters and latent quantities indexed by i and j, where $\epsilon_{1ij}, \epsilon_{2ij}, \theta_{1ij}, \theta_{2ij}$ and η_{ij}

are intensity function parameters (see Chapter II). for details. Given the study type $T_i = j$ and all other latent quantities denoted by "$\bullet$", we have the following BHICP model,

$$\begin{aligned} \mathbf{X}_i \mid T_i = j, \bullet &\overset{i.i.d.}{\sim} \text{Poisson}[\mathcal{B}, \lambda_{1j}(x \mid \mathbf{z}_j, \mathbf{\Sigma}_j, \Theta_{ij})] \\ \mathbf{Y}_{ij} \mid \bullet &\overset{i.i.d.}{\sim} \text{Poisson}[\mathcal{B}, \lambda_{2j}(y \mid \mathbf{z}_j, \mathbf{\Sigma}_j, \Theta_{ij})] \\ \mathbf{Z}_j \mid \beta_j &\overset{i.i.d.}{\sim} \text{Poisson}[\mathcal{B}, \beta_j] \end{aligned}$$

where the intensity functions λ_{1j} and λ_{2j} are given by:

$$\begin{aligned} \lambda_{1j}(x \mid \mathbf{z}_j, \mathbf{\Sigma}_j, \Theta_{ij}) &= \epsilon_{1ij} + \sum_{z \in \mathbf{z}_j} \theta_{1ij}\phi(x; z, \Sigma_z) + \sum_{y \in \mathbf{y}_{ij}} \eta_{ij}\phi(x; y, \Psi_y), \\ \lambda_{2j}(y \mid \mathbf{z}_j, \mathbf{\Sigma}_j, \Theta_{ij}) &= \epsilon_{2ij} + \sum_{z \in \mathbf{z}_j} \theta_{2ij}\phi(y; z, \Sigma_z). \end{aligned}$$

Let $\Theta_j = \{\mathbf{z}_j, \mathbf{\Sigma}_j, \Theta_{(n+1)j}\}$ and let $\mathbf{x}^j = \{\mathbf{x}_i : T_i = j\}$ denote all the foci that are reported from studies of type j. We have the follow corollary of Proposition 1 for the BHICP model.

Corollary 1. *In the BHICP model, the posterior predictive distribution* (4.4) *for* T_{n+1} *is proportional to*

$$p_j \int \exp\left\{-\int_{\mathcal{B}} \lambda_{1j}(x \mid \Theta_j)dx\right\}\left\{\prod_{x \in \mathbf{x}_{n+1}} \lambda_{1j}(x \mid \Theta_j)\right\}\pi(\Theta_j \mid \mathbf{x}^j)d\Theta_j, \tag{4.14}$$

where

$$\begin{aligned} \pi(\Theta_j \mid \mathbf{x}^j) &= \pi(\mathbf{y}_{(n+1)j} \mid \mathbf{z}_j, \mathbf{\Sigma}_j, \mathbf{\Psi}_{(n+1)j}, \epsilon_{2(n+1)j}, \theta_{2(n+1)j})\pi(\mathbf{z}_j, \mathbf{\Sigma}_j \mid \mathbf{x}^j) \\ &\times \pi(\mathbf{\Psi}_{(n+1)j})\pi(\epsilon_{1(n+1)j})\pi(\theta_{1(n+1)j})\pi(\epsilon_{2(n+1)j}), \pi(\theta_{2(n+1)j})\pi(\eta_{(n+1)j}). \end{aligned} \tag{4.15}$$

This implies that to draw Θ_j given $\mathbf{x}^j$, we first draw $(\mathbf{z}_j, \mathbf{\Sigma}_j)$ given $\mathbf{x}^j$ for $j = 1, \ldots, J$ by applying the BHICP model running the posterior simulations J times, one for each study type j (see Chapter II for more details). Then draw $\epsilon_{1(n+1)j}$, $\theta_{1(n+1)j}$, $\epsilon_{2(n+1)j}$, $\theta_{2(n+1)j}$, $\eta_{(n+1)j}$ and $\mathbf{\Psi}_{(n+1)j}$ from their respective prior distribution. And finally draw $\mathbf{y}_{(n+1)j}$ given $\epsilon_{2(n+1)j}$, $\theta_{2(n+1)j}$, $\mathbf{\Psi}_{(n+1)j}$, $\mathbf{z}_j$ and $\mathbf{\Sigma}_j$ from (4.14).

4.2.3 The HPGRF Model

For each type j, the foci reported from study i, i.e. $\mathbf{x}_i$, are one realization of a Poisson point process with random intensity λ_j, where λ_j is modeled as a spatial convolution of a kernel function $k_{\sigma_j^2}(x, y)$ and a gamma random field $G_j(dy)$ with base measure $G_0(dy)$ and scale parameter τ. To induce correlation between studies, we further assume that $G_0(dy)$ is also a gamma random field with base measure $\alpha(dy)$ and scale parameter β. In summary, we have that

$$\begin{aligned}
\mathbf{X}_i \mid T_i = j, \bullet &\overset{i.i.d.}{\sim} \text{Poisson}[\mathcal{B}, \lambda_j(x \mid G_j, \sigma_j^2)], \\
\lambda_j(x \mid G_j, \sigma_j^2) &= \int_{\mathcal{B}} k_{\sigma_j^2}(x, y) G_j(dy), \\
G_j(dy) &\overset{i.i.d,}{\sim} \mathcal{GRF}(G_0(dy), \tau), \\
G_0(dy) &\sim \mathcal{GRF}(\alpha(dy), \beta). \qquad (4.16)
\end{aligned}$$

Let $\Theta = \{\{\sigma_j^2\}_{j=1}^J, \{G_j)\}_{j=0}^J\}$. Then, based on the posterior simulation algorithm proposed in Chapter III, propositions 1 and 2 allow us to perform reverse inference and produce the LOOCV classification rates.

4.3 Single Focus Impact Map

To understand the contribution of an individual focus to the classification, we consider a voxel wise “Single Focus Impact Map” (SFIM). Define $S_j(\xi)$ as the expected change in the predictive probability of study type $T_{n+1} = j$ given $\mathbf{x}_{n+1}$ and $\mathcal{D}_n$ when adding a focus ξ at voxel v to $\mathbf{x}_{n+1}$. The expectation is taken with respect to the predictive distribution of $\mathbf{X}_{n+1}$ given $\mathcal{D}_n$. i.e. For $\xi \in v \subset \mathcal{B}$ and each type j,

$$\begin{aligned}
&S_j(\xi) := \\
&E_{\mathbf{X}_{n+1}|\mathcal{D}_n}\left[\Pr(T_{n+1} = j \mid \mathbf{x}_{n+1} \cup \{\xi\}, \mathcal{D}_n) - \Pr(T_{n+1} = j \mid \mathbf{x}_{n+1}, \mathcal{D}_n)\right] \qquad (4.17)
\end{aligned}$$

The following proposition allows easy computation of $S_j(\xi)$.

Proposition 3. *The SFIM map defined by* (4.18) *is given by*

$$S_j(\xi) = E_{\mathbf{X}_{n+1},\Theta|\mathcal{D}_n}\left[\frac{p_j\lambda_j(\xi \mid \Theta)\pi(\mathbf{x}_{n+1} \mid T_{n+1}=j,\Theta)}{\sum_{j'=1}^{J} p_{j'}\lambda_{j'}(\xi \mid \Theta)\pi(\mathbf{x}_{n+1} \mid T_{n+1}=j',\Theta)}\right] - p_j. \tag{4.18}$$

Thus, $S_j(\xi) \in [-p_j, 1-p_j]$.

This implies that $S_j(\xi)$ is positive when a type j focus, ξ, uniquely occurs at voxel v. If only a type j focus, ξ, is absent at voxel v, then $S_j(\xi)$ will be negative.

The algorithm for SFIM:

- **Input**: The observed data $\mathcal{D}_n$ and the total number of simulations K.
- **Step 1**: Run a Bayesian spatial point process model to obtain posterior draws $\Theta^{(k)} \sim \pi(\Theta \mid \mathcal{D}_n)$, for $k = 1, \ldots, K$.
- **Step 2**: Draw $t_{n+1}^{(k)}$ from the prior probabilities $\{p_j\}_{i=1}^{J}$, for $k = 1, \ldots, K$.
- **Step 3**: Draw $\mathbf{x}_{n+1}^{(k)}$ from the density $\pi(\mathbf{x}_{n+1} \mid T_{n+1} = t_{n+1}^{(k)}, \Theta^{(k)})$, for $k = 1, \ldots, K$.
- **Step 4**: Let $\{v_1, \ldots, v_N\}$ denote all the voxels in $\mathcal{B}$, and let ξ_i be the center of v_i. For $i = 1, \ldots, N$, $j = 1, \ldots, J$ and $k = 1, \ldots, K$, compute

$$L_j^{(k)}(\xi_i) = \lambda_j(\xi_i \mid \Theta^{(k)})\pi(\mathbf{x}_{n+1}^{(k)} \mid T_{n+1} = j, \Theta^{(k)}). \tag{4.19}$$

- **Output**: The salience map $S_j(\xi_i)$, for $i = 1, \ldots, N$ and $j = 1, \ldots, J$, is estimated by

$$\widehat{S}_j(\xi_i) = \frac{1}{K}\sum_{k=1}^{K} \frac{p_j L_j^{(k)}(\xi_i)}{\sum_{j'=1}^{J} p_{j'} L_{j'}^{(k)}(\xi_i)} - p_j. \tag{4.20}$$

4.4 Application

To demonstrate our method, we use the meta analysis of emotions previously discussed in Chapters II and III. We only consider the five emotions anger, disgust, fear, happy and sad, leaving 219 studies. Table 4.1 shows summary statistics for each type of emotion, where anger has the least number of studies and foci, fear has the most number of studies and foci, and sad has the highest average number of foci per study. We assume equal prior probability, $p_j = 0.2$, for each type, and fit our Bayesian spatial point process classifier based on the BHICP model and the HPGRF model. Table 4.2 shows the LOOCV classifications rates based on our BHICP and HPGRF models as well as those based on MKDA using the NBC. Our spatial classifier correctly classifies 181 (BHICP) and 186 (HPGRF) of the 219 studies, for overall rates of 0.83 (BHICP) and 0.85 (HPGRF), far above random chance of 0.20. Averaging correct classification rates over emotions which gives equal weight to all emotions, provides an average rates of 0.82 and 0.85 respectively for the BHICP and HPGRF based spatial classifier. The MKDA based NBC (kernel radius is 10 mm) correctly classifies 99 studies with an overall rate of 0.45 and an average rate of 0.36. Changing the MKDA kernel radius to 5mm, 15mm and 20mm faired no better. The single focus impact maps are displayed in Figure 4.1 along with prediction intensity maps and population center intensity maps. On the axial slice located at Z $= -20$mm, the right amygdala (pointed to by white arrows) is visible in both study level intensity maps and population center intensity maps. There is greater spatial precision in the population center map. Since most emotions have some right amygdala activation, inclusion of a new focus in or near the right amygdala does not result in strong signal in the SFIM. However anger, which shows evidence of right lateral orbito-frontal activation (pointed to by red arrows), is unique and thus appears as strong positive signal in the SFIM map. On the axial slice, Z = 26mm, the predictive intensity of study foci and the population center posterior intensity reveals that sad, happy and disgust have some precuneus activation (white arrows), except anger. The SFIM reflects this, with anger studies having a negative predictive value for precuneus activations: more precisely, for a random study, if you add one focus in the precuneus, the predictive probability of anger decreases, increasing the predictive probability of disgust and sad study types. The SFIM for anger also shows positive predictive value of activations in the inferior frontal gyrus (pointed to by red arrows) relative to other emotion types. The peak intensity in the anger SFIM is 0.024, indicating that the addition of a single focus to the inferior frontal gyrus increases the predictive probability 0.024.

4.5 Simulation Studies

In this section, we conduct simulation studies to evaluate the performance of our methods. We compared the prediction accuracy of the BHICP, the HPGRF and the NBC methods for different sets of simulated data.

4.5.1 Simulated Data Sets

We consider four types of foci and assume each type has 100 studies. Specifically, we generate the type indicator $t_i = j$, if $j - 1 \leq i/100 < j$, for $i = 1, \ldots, 400$. Given $\{t_i\}_{i=1}^{400}$, we simulate foci $\mathbf{x}_i$ from Poisson cluster processes on brain $\mathcal{B}$ (MNI space) with mixture intensity functions:

$$[\mathbf{X}_i \mid \boldsymbol{T}_i = j, \lambda_j] \sim \text{Poisson}[\mathcal{B}, \lambda_j(x)], \text{ for } j = 1, \ldots, 4. \tag{4.21}$$

where $\lambda_j(x) = \epsilon + \sum_{l=1}^{3} \theta_{jl} \phi(x \mid \mu_l, \Sigma_l)$. The locations of cluster centers are given by $\mu_1 = (-30, 20, 15)'$, $\mu_2 = (30, -10, -10)'$ and $\mu_3 = (0, -40, 40)'$. The cluster shapes are given by $\Sigma_1 = \begin{pmatrix} 40 & 1 & 1 \\ 1 & 40 & 1 \\ 1 & 1 & 40 \end{pmatrix}$, $\Sigma_2 = \begin{pmatrix} 30 & 4 & 4 \\ 4 & 30 & 4 \\ 4 & 4 & 30 \end{pmatrix}$, $\Sigma_3 = \begin{pmatrix} 20 & 2 & 2 \\ 2 & 20 & 2 \\ 2 & 2 & 20 \end{pmatrix}$. The background intensity is $\epsilon = 1/|\mathcal{B}|$. For the choices of cluster intensities $\{\{\theta_{jl}\}_{l=1}^{3}\}_{j=1}^{4}$, we consider the following two cases:

Case 1

$$\theta_{11} = 1, \theta_{12} = 1, \theta_{13} = 0,$$

$$\theta_{21} = 0, \theta_{22} = 1, \theta_{23} = 1,$$

$$\theta_{31} = 1, \theta_{32} = 0, \theta_{33} = 1,$$

$$\theta_{41} = 1, \theta_{42} = 1, \theta_{43} = 1.$$

Case 2

$$\theta_{11} = 1, \theta_{12} = 1, \theta_{13} = 1,$$

$$\theta_{21} = 1, \theta_{22} = 1, \theta_{23} = 1,$$

$$\theta_{31} = 2, \theta_{32} = 2, \theta_{33} = 2,$$

$$\theta_{41} = 2, \theta_{42} = 2, \theta_{43} = 2.$$

Figure 4.2 shows the simulated foci for the two different cases. In case 1, type 1 and type 2 share clustering in region 1, type 2 and type 3 share clustering in region 3, and type 1 and type 3 share clustering in region 2. Type 4 shows clustering in all three regions. In case 2, all four types share clustering in the same region but with different intensities. Type 1 and type 2 have the same intensity. Type 3 and type 4 have identical intensities. The cluster intensities of type 3 and type 4 aer twice as large as those of type 1 and type 2.

4.5.2 Simulation Performance

We apply our Bayesian spatial point process classifier and the MKDA based NBC method to the simulated data. We assume equal prior probability of each type in the classification. For the BHICP model of each type, we specify the priors according to Chapter II, where the expected number of population centers is 3. In the HPGRF model we use the same priors that are specified by Chapter III. We compute the LOOCV classification rates for the BHICP model, the HPGRF model and the MKDA based NBC method for the two cases (Tables 4.3 and 4.4). For both two cases, the BHICP and the HPGRF models provide better prediction accuracy than the NBC method. The overall BHICP classification accuracy is comparable to that of the HPGRF model in both cases. In case 1, both the BHCIP and the HPGRF models have overall classification rates greater than 0.75, much larger than the NBC correct classification rate of 0.575. The correct classification rates for types 1, 2 and 3 are all higher than that of type 4 for all three models. This is due to the fact that type 4 shares clustering with other types. In case 2, none of the models can distinguish type 1 and type 2, because both of them are simulated from the same intensity, which is also true for type 3 and type 4. However, in the BHICP and the HPGRF confusion matrices in Table 4.4, only a few studies of type 1/type 2 are misclassified as type 3/type 4. The models can detect the differences in the absolute number between the types of foci that show clustering in the same region. The NBC method provides almost the same misclassification rates for each type.

4.6 Discussion

We propose a Bayesian spatial point process classifier for reverse inference for neuroimaging meta analysis. Our method obtains accuracy rates twice those of the naive Bayesian classifier. Possible reasons for this superior performance include: 1) We use an explicit spatial point process model that better captures the spatial structure of foci than the mass-univariate NBC. Our approach jointly models the random number and locations of foci, while the NBC ignores the spatial dependence between voxels. 2) Our hierarchical spatial model is a more accurate representation of the true data generating mechanism, relative to the MKDA kernel mapping of points to voxel-wise images. 3) Our fully Bayesian model captures more sources of variation, and appropriately conveys the certainty (or lack there of) in the computation of the predictive probabilities that determine the classification outcome. Finally these results provide evidence against a homogeneous representation of different emotions in the brain.

In the future, one extension of the BHCIP model is to build up a joint cluster process model for multiple types of point patterns, where their intensity functions are assumed to be driven by the same population center process, but with distinct parameters. This would enable the detection of common population centers shared by some, or all, sub types, as well as distinct population centers for each sub type. A joint model of this type would be more parsimonious than the current BHCIP model and thus should increases statistical efficiency.

Table 4.1: Summary statistics from the emotion meta analysis data set

Statistics	Sad	Happy	Anger	Fear	Disgust
Number of studies	45	36	26	68	44
Number of foci	346	177	166	367	337
Avg. number of foci per study	7.69	4.92	6.38	5.40	7.66

Table 4.2: The LOOCV classification rates

Naive Bayes Classifier (overall: 0.45, average: 0.37)

Truth	Classification Rates				
	sad	happy	anger	fear	disgust
sad	0.38	0.11	0.07	0.40	0.04
happy	0.11	0.25	0.03	0.56	0.06
anger	0.12	0.23	0.00	0.50	0.15
fear	0.06	0.06	0.01	0.81	0.06
disgust	0.09	0.16	0.05	0.32	0.39

Bayesian Hierarchical Independent Cluster Process Model
(overall: 0.83, average: 0.82)

Truth	Classification Rates				
	sad	happy	anger	fear	disgust
sad	0.78	0.00	0.11	0.04	0.07
happy	0.06	0.92	0.00	0.03	0.00
anger	0.08	0.08	0.69	0.15	0.00
fear	0.13	0.01	0.00	0.85	0.00
disgust	0.05	0.02	0.02	0.07	0.84

Hierarchical Poisson/Gamma Random Fields Model
(overall: 0.85, average: 0.85)

Truth	Classification Rates				
	sad	happy	anger	fear	disgust
sad	0.91	0.02	0.00	0.07	0.00
happy	0.03	0.83	0.03	0.08	0.03
anger	0.00	0.08	0.77	0.15	0.00
fear	0.01	0.07	0.06	0.81	0.04
disgust	0.02	0.02	0.00	0.02	0.93

Table 4.3: Simulation study results: Case 1
Naive Bayes Classifier (overall: 0.575)

Truth	Classification Rates			
	Type 1	Type 2	Type 3	Type 4
Type 1	0.61	0.16	0.16	0.07
Type 2	0.09	0.83	0.04	0.04
Type 3	0.19	0.19	0.57	0.05
Type 4	0.18	0.26	0.27	0.29

Bayesian Hierarchical Independent Cluster Process Model
(overall: 0.77)

Truth	Classification Rates			
	Type 1	Type 2	Type 3	Type 4
Type 1	0.81	0.06	0.09	0.04
Type 2	0.05	0.84	0.06	0.05
Type 3	0.10	0.05	0.79	0.06
Type 4	0.12	0.10	0.13	0.65

Hierarchical Poisson/Gamma Random Fields Model
(overall: 0.76)

Truth	Classification Rates			
	Type 1	Type 2	Type 3	Type 4
Type 1	0.78	0.07	0.07	0.08
Type 2	0.07	0.83	0.05	0.05
Type 3	0.10	0.06	0.78	0.06
Type 4	0.10	0.13	0.13	0.64

Table 4.4: Simulation study results: Case 2
Naive Bayes Classifier (overall: 0.32)

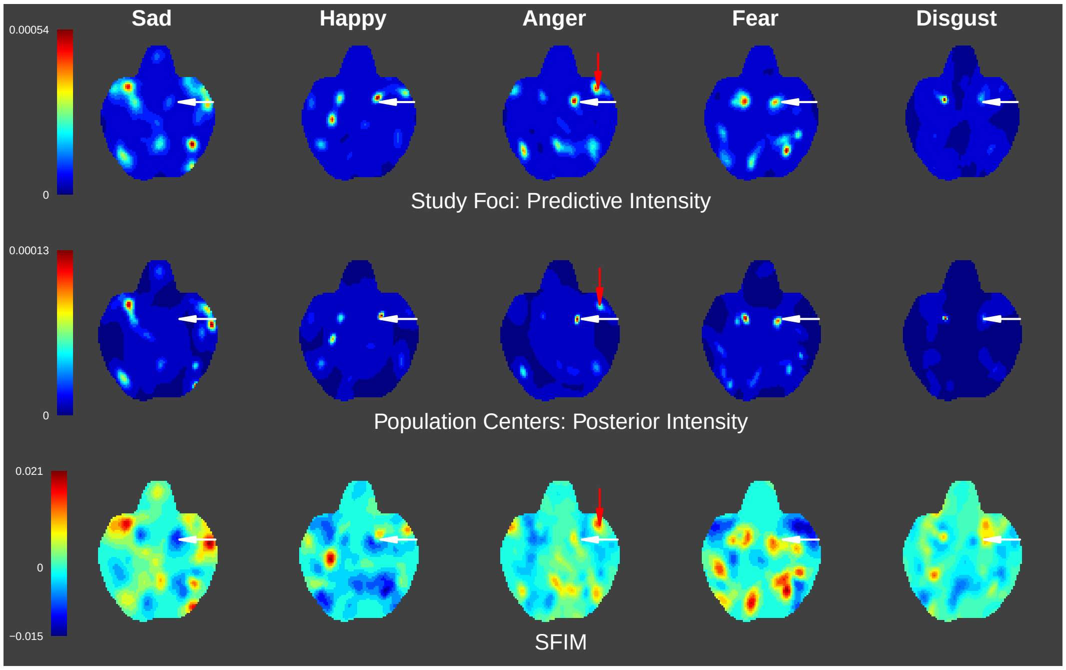
Truth	Classification Rates			
	Type 1	Type 2	Type 3	Type 4
Type 1	0.35	0.33	0.12	0.20
Type 2	0.42	0.31	0.10	0.17
Type 3	0.16	0.13	0.31	0.40
Type 4	0.22	0.10	0.38	0.30

Bayesian Hierarchical Independent Cluster Process Model
(overall: 0.45)

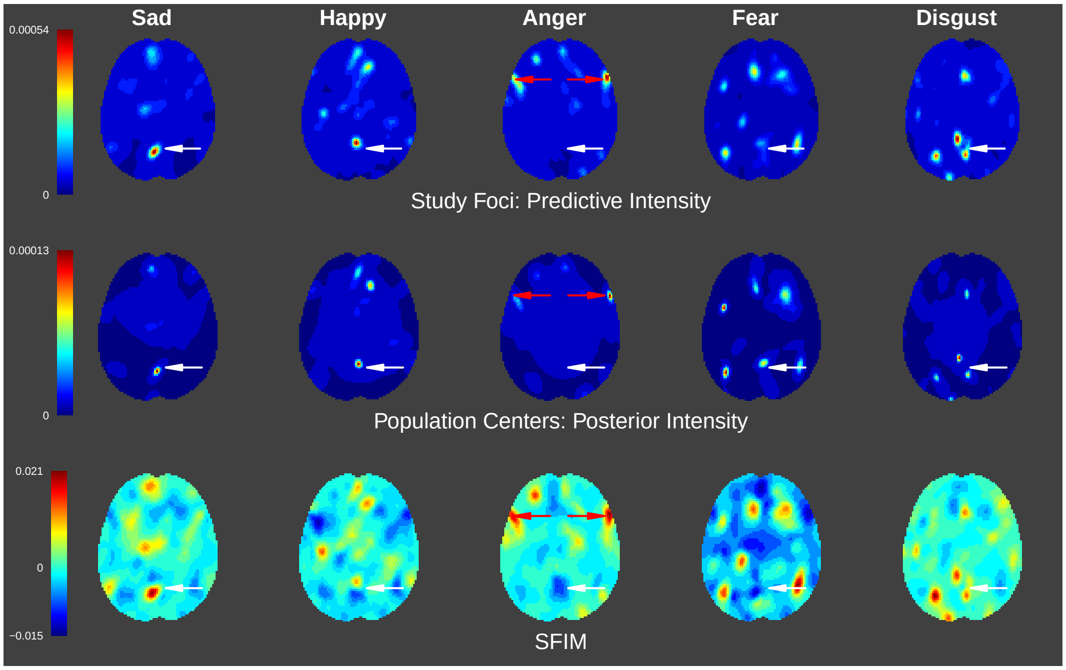
Truth	Classification Rates			
	Type 1	Type 2	Type 3	Type 4
Type 1	0.51	0.40	0.02	0.07
Type 2	0.46	0.45	0.05	0.04
Type 3	0.03	0.06	0.43	0.48
Type 4	0.01	0.06	0.41	0.42

Hierarchical Poisson/Gamma Random Fields Model
(overall: 0.51)

Truth	Classification Rates			
	Type 1	Type 2	Type 3	Type 4
Type 1	0.47	0.49	0.01	0.03
Type 2	0.37	0.54	0.04	0.05
Type 3	0.04	0.07	0.49	0.40
Type 4	0.02	0.02	0.42	0.54

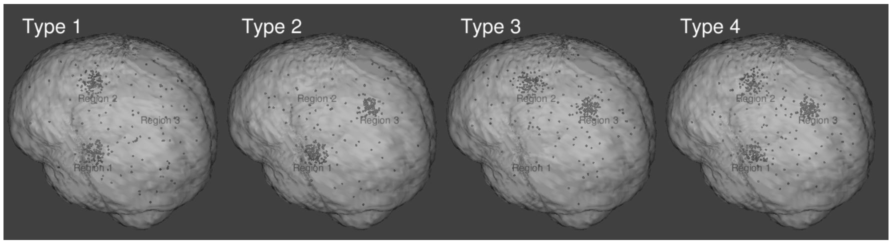


$Z = -20mm$

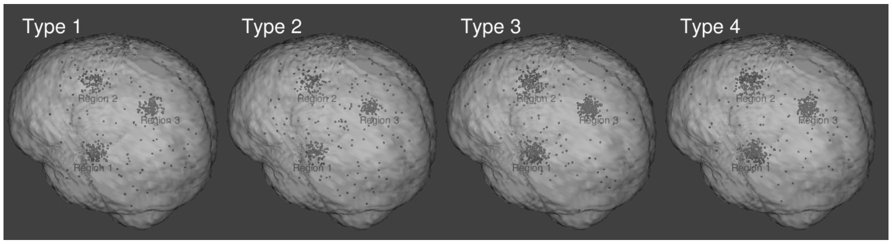


$Z = 26mm$

Figure 4.1: Two axial slices (Z = -20mm, 26mm) of the predictive intensity of study foci, the posterior intensity of population centers, and single focus impact maps produced by BHICP model for the meta analysis of five emotions.



Case 1



Case 2

Figure 4.2: Simulated Data.

CHAPTER V

Conclusion

Motivated by a meta analysis of 437 emotion studies, in this dissertation, we develop a series of Bayesian hierarchical spatial point process models for functional neuroimaging meta analysis and related problems.

In Chapter II, we produce a fully Bayesian hierarchical spatial model, i.e. the BHICP model, for neuroimaging meta analysis. We explicitly model the clustering of study foci about latent activation centers (accounting for multiple peaks within a study describing a single anatomical region), which in turn cluster about latent population centers. In particular, our model dissociates between inter-study spread of foci, and spatial uncertainty in population response location. Furthermore, since we adopt a Bayesian modeling approach, other quantitative information can be extracted from our model that cannot be deduced from current methods.

In the future, one extension of the BHICP model is to incorporate multiple sources of information as covariates into the study level intensity functions such as sample size, nominal significance level, and whether or not the study is adjusted for multiple comparisons. The covariate adjusted model can increase the power of detection of population centers and reduce bias. We would also like to study the identifiability of the hierarchical independent cluster process model, particularly focusing on the assessment of Bayesian learning of parameters, such as location and variability of population centers.

In Chapter III, we develop a hierarchical Poisson/Gamma random field model for multi type spatial point processes. We have shown that our model not only estimates the intensity function more accurately for each type of point pattern, but also provides a “population mean” intensity which can be used to identify common regions of clustering shared between the types. Although the development of this model is motivated by the analysis of multi type emotion meta analysis, it can be applied to any problem that involves multi type spatial point patterns.

In the future, we plan to extend the model using multiple levels of hierarchy where the number of levels and the structure of the hierarchy depend on the needs of the data analysis. This model is particularly useful for complex data analysis. In the meta analysis of emotion, for example, based on prior knowledge, some regions in the brain are activated for both anger and fear, but not for other emotions. A multi-level hierarchical extension of our model could identify these regions.

The current algorithm is based on the Lévy measure construction which involves a large number of parameters and must be greater than the number of observed points. This results in an expensive computational cost. Thus, another interesting future direction is to develop a more efficient algorithm for the posterior computation of our model. One potential solution is to develop a equivalent model representation with a random number of parameters in a similar fashion to the Chinese Restaurant Franchise (Teh et al. 2006) for hierarchical Dirichlet process.

In Chapter IV, we focus on reverse inference based on the meta analysis of functional neuroimaging data, for which we propose a Bayesian spatial point process classifier for multiple types of point patterns. With appropriate modeling of the intensity function for each type of point pattern, we have shown that a classifier can be constructed based on the posterior probability of the type, given a particular point pattern. This posterior probability, for a new study, can be efficiently estimated via an importance sampling approach. We also propose a statistical map as an exploratory tool to discover the impact of brain regions on distinguishing cognitive states. The proposed method is illustrated by a meta analysis of 219 emotion studies with five types. This model shows a much higher correct classification rate compared with a naive Bayes classifier, and the proposed statistical map provides valuable insight into the relationships between emotions and brain activity. Our modeling framework can integrate any model that has explicit estimates of the intensity function or density function for each type of point pattern. We have demonstrated that the BHICP and the HPGRF models can be adapted for a classifier through our methods. We have found that the HPGRF model which jointly analyses the multi type spatial point patterns results in higher accuracy than the BHICP model that independently fits each type of point pattern. However, the BHICP model has the advantage that it dissociates between inter-study spread of foci and the spatial uncertainty in population response locations. This motivates future work to build a joint cluster process model for multi type point patterns, where their intensity functions are assumed to be driven by a common population center process, but with distinct parameters. For a particular type, the cluster intensity can be very small or close to zero, indicating that it does not cluster about this population center. This enables detection of the common population centers shared by some or all of the types of point pattern as well as the particular population centers for each type. This joint model is more parsimonious than the current model and thus would increase the statistical efficiency. Furthermore, it may produce higher correct classification rates.

Another promising future direction for the analysis of spatial point patterns is to develop models for Bayesian random shape spatial clustering. The Matérn cluster process makes an assumption that points cluster in spheres. In some applications, spatial point patterns may cluster in configurations that are not well represented by a sphere. Overlapping spherical clusters may be able to capture some features of these clusters, however, the number and shape of unique clusters may also be of interest. Thus, the development of a Bayesian random shape spatial point process model that generalizes the Matérn cluster process in 3D is of intereset. We can represent the shape of clusters as closed surfaces using 3D spherical coordinates and build up a stochastic model for the radial function. The key idea is to utilize the spectral representation of a homogeneous and isotropic Gaussian random field on the unit sphere (Yadrenko 1983). We obtain the random radial function of the shapes by exponentiating this process. Simulation studies show the model has good performance on fitting different shapes of cluster point patterns.

In the future, we plan to apply this model to the meta analysis of functional neuroimaging and compare it with other models. The general idea may also be applied in other random shape analyses. For example, we can incorporate the random shape framework into a spatial-mixture model (Woolrich et al. 2005) that segments the functional statistical parametric maps into three classes: activation, non-activation and de-activation. The random shapes would represent the boundaries of the regions for each class, which specifies the prior model of the spatial class labels. It avoids estimating the normalizing constant compared with the regular discrete Markov random fields models.

APPENDICES

APPENDIX A

The Details of Bayesian Hierarchical Independent Cluster Process Model

A.1 Probability Equivalence Assumption

The *probability equivalence assumption*: the probability that a type 0 focus in study c clusters about a population center $z \in \mathbf{z}$ and the probability that a study center in study c clusters about the same population center are equal over all studies, for $c = 1, \dots, C$. Let x, y be two points in the brain $\mathcal{B}$, then we have

$$\begin{aligned}
\Pr(x \in \mathbf{X}^0_{cz} \mid x \in \mathbf{X}^0_c) &= \Pr(y \in \mathbf{Y}_{cz} \mid y \in \mathbf{Y}_c) \\
\frac{\Pr(x \in \mathbf{X}^0_{cz})}{\Pr(x \in \mathbf{X}^0_c)} &= \frac{\Pr(y \in \mathbf{Y}_{cz})}{\Pr(y \in \mathbf{Y}_c)} \\
\frac{E[I(x \in \mathbf{X}^0_{cz})]}{E[I(x \in \mathbf{X}^0_c)]} &= \frac{E[I(y \in \mathbf{Y}_{cz})]}{E[I(y \in \mathbf{Y}_c)]} \qquad \text{(A.1)} \\
\frac{E[N_{\mathbf{X}^0_{cz}}(\mathcal{B})]}{E[N_{\mathbf{X}^0_c}(\mathcal{B})]} &= \frac{E[N_{\mathbf{Y}_{cz}}(\mathcal{B})]}{E[N_{\mathbf{Y}_c}(\mathcal{B})]}, \qquad \text{(A.2)}
\end{aligned}$$

where (A.2) follows from (A.1) by the following argument. Suppose there are two point processes $\mathbf{A}, \mathbf{B} \subset \mathcal{S}$ and $\mathbf{B} \neq \emptyset$, for any two points $\xi, \eta \in \mathcal{S}$, then $\Pr(\xi \in \mathbf{A}) = \Pr(\eta \in \mathbf{A})$ and $\Pr(\xi \in \mathbf{B}) = \Pr(\eta \in \mathbf{B}) > 0$, thus,

$$\frac{\Pr(\xi \in \mathbf{A})}{\Pr(\xi \in \mathbf{B})} = \frac{\Pr(\eta \in \mathbf{A})}{\Pr(\eta \in \mathbf{B})} \Rightarrow \frac{E[I(\xi \in \mathbf{A})]}{E[I(\xi \in \mathbf{B})]} = \frac{E[I(\eta \in \mathbf{A})]}{E[I(\eta \in \mathbf{B})]} \equiv R.$$

Therefore,

$$\begin{aligned}\frac{E[N_{\mathbf{A}}(\mathcal{S})]}{E[N_{\mathbf{B}}(\mathcal{S})]} &= \frac{\sum_{\eta\in\mathcal{S}} E[I(\eta\in\mathbf{A})]}{\sum_{\eta\in\mathcal{S}} E[I(\eta\in\mathbf{B})]} \\ &= \frac{\sum_{\eta\in\mathcal{S}} R\cdot E[I(\eta\in\mathbf{B})]}{\sum_{\eta\in\mathcal{S}} E[I(\eta\in\mathbf{B})]} = R = \frac{E[I(\xi\in\mathbf{A})]}{E[I(\xi\in\mathbf{B})]}, \quad \text{for any } \xi\in\mathcal{S},\end{aligned}$$

where $N_\bullet(\mathcal{S}) = \sum_{a\in\mathcal{S}} I(a\in\bullet)$.

According to the definitions of $\mathbf{X}_c^0$, $\mathbf{X}_{cz}^0$, $\mathbf{Y}_c$ and $\mathbf{Y}_{cz}$ and the intensity function (3) and (6) in Section 2.2, (A.2) implies, for $c = 1, \ldots, C$,

$$\text{(A.3)}\quad p_z \equiv \frac{\theta_{1c}\Phi(\mathcal{B}; z, \mathbf{\Sigma}_z)}{\epsilon_{1c}|\mathcal{B}| + \sum_{z\in\mathbf{z}}\theta_{1c}\Phi(\mathcal{B}; z, \mathbf{\Sigma}_z)} \quad = \quad \frac{\theta_{2c}\Phi(\mathcal{B}; z, \mathbf{\Sigma}_z)}{\epsilon_{2c}|\mathcal{B}| + \sum_{z\in\mathbf{z}}\theta_{2c}\Phi(\mathcal{B}; z, \mathbf{\Sigma}_z)}.$$

By routine calculations, we have

$$\text{(A.4)}\qquad \frac{\theta_{1c}}{\epsilon_{1c}} = \frac{\theta_{2c}}{\epsilon_{2c}} = \frac{1}{|\mathcal{B}|}\left(\frac{\Phi(\mathcal{B}; z, \mathbf{\Sigma}_z)}{p_z} - \sum_{z'\in\mathbf{z}}\Phi(\mathcal{B}; z', \mathbf{\Sigma}'_z)\right) \equiv q_z.$$

From the definition of θ and ϵ in equation (8) in Section 2.2, we have

$$\text{(A.5)}\qquad \frac{\theta}{\epsilon} = \frac{\sum_{c=1}^{C}(\theta_{1c}+\theta_{2c})}{\sum_{c=1}^{C}(\epsilon_{1c}+\epsilon_{2c})} = \frac{\sum_{c=1}^{C} q_z(\epsilon_{1c}+\epsilon_{2c})}{\sum_{c=1}^{C}(\epsilon_{1c}+\epsilon_{2c})} = q_z = \frac{\theta_{1c}}{\epsilon_{1c}} = \frac{\theta_{2c}}{\epsilon_{2c}}.$$

This further implies that we can define ρ_{1c} and ρ_{2c} as follows:

$$\text{(A.6)}\qquad \rho_{1c} \equiv \frac{\epsilon_{1c}}{\epsilon} = \frac{\theta_{1c}}{\theta}, \qquad \rho_{2c} \equiv \frac{\epsilon_{2c}}{\epsilon} = \frac{\theta_{2c}}{\theta},$$

such that

$$\text{(A.7)}\qquad \sum_{c=1}^{C}(\rho_{1c}+\rho_{2c}) = \frac{\sum_{c=1}^{C}(\theta_{1c}+\theta_{2c})}{\theta} = \frac{\sum_{c=1}^{C}(\epsilon_{1c}+\epsilon_{2c})}{\epsilon} = 1.$$

Write $\boldsymbol{\rho} = (\rho_{11}, \cdots, \rho_{1C}, \rho_{21}, \cdots, \rho_{2C})$. The parameter set $\{\epsilon_{1c}, \theta_{1c}, \epsilon_{2c}, \theta_{2c}\}_{c=1}^{C}$ can be now be reparametrized by $\{\boldsymbol{\rho}, \theta, \epsilon\}$. Thus, the number of parameters is reduced by $2C-1$.

A.2 Algorithm Details

In this section, we provide algorithm details and pseudo code. We first present the posterior distribution, and then discuss the details of the continuous time spatial birth-and-death processes for simulating from the posterior study activation center processes and the posterior population center process. We also provide details on updating all parameters in the hybrid MCMC algorithm. We then discuss some issues on Normal and Student-t probability computations, show how to estimate the posterior intensity function for the population center process and the 95% credible ellipsoids.

A.2.1 Posterior Distribution

The joint posterior of model parameters, including latent variables, is

$$\begin{aligned}
&\pi[\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_C, (\mathbf{y}, \boldsymbol{\Psi})_1, \ldots, (\mathbf{y}, \boldsymbol{\Psi})_C, (\mathbf{z}, \boldsymbol{\Sigma}), \boldsymbol{\epsilon}, \boldsymbol{\theta}, \boldsymbol{\eta}, \beta, \boldsymbol{T} \mid \mathbf{x}_1, \ldots \mathbf{x}_C] \\
&\quad \propto \prod_{c=1}^{C} \Big\{ \pi\left[(\mathbf{x}, \boldsymbol{\delta})_c \mid \epsilon_{1c}, \theta_{1c}, \eta_c, (\mathbf{z}, \boldsymbol{\Sigma}), (\mathbf{y}, \boldsymbol{\Psi})_c\right] \pi(\epsilon_{1c}) \pi(\eta_c) \pi(\theta_{1c}) \Big\} \times \\
&\quad \prod_{c=1}^{C} \Big\{ \pi\left[(\mathbf{y}, \boldsymbol{\Psi})_c \mid \epsilon_{2c}, \theta_{2c}, (\mathbf{z}, \boldsymbol{\Sigma})\right] \pi(\epsilon_{2c}) \pi(\theta_{2c}) \Big\} \times \pi\left[(\mathbf{z}, \boldsymbol{\Sigma}) \mid \beta, \boldsymbol{T}\right] \pi(\beta) \pi(\boldsymbol{T}),
\end{aligned} \tag{A.8}$$

where $\boldsymbol{\epsilon} = (\epsilon_{11}, \ldots, \epsilon_{1C}, \epsilon_{21}, \ldots, \epsilon_{2C})$, $\boldsymbol{\theta} = (\theta_{11}, \ldots, \theta_{1C}, \theta_{21}, \ldots, \theta_{2C})$ and $\boldsymbol{\eta} = (\eta_1 , \cdots , \eta_C)$. With the probability equivalence assumption, we have that $\epsilon_{dc} = \rho_{dc}\epsilon$ and $\theta_{dc} = \rho_{dc}\theta$ for $d = 1, 2$ and $c = 1, \ldots, C$, Thus, (A.8) is equivalent to

$$\begin{aligned}
&\pi[\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_C, (\mathbf{y}, \boldsymbol{\Psi})_1, \ldots, (\mathbf{y}, \boldsymbol{\Psi})_C, (\mathbf{z}, \boldsymbol{\Sigma}), \boldsymbol{\rho}, \boldsymbol{\eta}, \beta, \boldsymbol{T} \mid \mathbf{x}_1, \ldots \mathbf{x}_C] \\
&\quad \propto \prod_{c=1}^{C} \Big\{ \pi\left[(\mathbf{x}, \boldsymbol{\delta})_c \mid \rho_{1c}, \eta_c, (\mathbf{z}, \boldsymbol{\Sigma}), (\mathbf{y}, \boldsymbol{\Psi})_c\right] \pi(\eta_c) \Big\} \times \\
&\quad \prod_{c=1}^{C} \Big\{ \pi\left[(\mathbf{y}, \boldsymbol{\Psi})_c \mid \rho_{2c}, (\mathbf{z}, \boldsymbol{\Sigma})\right] \Big\} \pi(\boldsymbol{\rho}) \times \pi\left[(\mathbf{z}, \boldsymbol{\Sigma}) \mid \beta, \boldsymbol{T}\right] \pi(\beta) \pi(\boldsymbol{T}).
\end{aligned} \tag{A.9}$$

This is proportional to

$$\begin{aligned}
&\prod_{c=1}^{C}\left\{\pi\left[\mathbf{x}_c^0 \mid \rho_{1c}, (\mathbf{z}, \mathbf{\Sigma})\right] \prod_{x\in\mathbf{x}_c^0} \pi(\delta_x = 0)\right\} \times \\
&\prod_{c=1}^{C}\left\{\pi\left[\mathbf{x}_c^1 \mid \eta_c, (\mathbf{y}, \mathbf{\Psi})_c\right] \prod_{x\in\mathbf{x}_c^1} \pi(\delta_x = 1)\pi(\eta_c)\right\} \times \\
&\prod_{c=1}^{C}\left\{\pi\left[(\mathbf{y}, \mathbf{\Psi})_c \mid \rho_{2c}, (\mathbf{z}, \mathbf{\Sigma})\right]\right\}\pi(\boldsymbol{\rho}) \times \pi\left[(\mathbf{z}, \mathbf{\Sigma}) \mid \beta, \boldsymbol{T}\right]\pi(\beta)\pi(\boldsymbol{T}). \qquad \text{(A.10)}
\end{aligned}$$

We propose an hybrid algorithm with two continuous time spatial birth-and-death processes embedded in a standard MCMC algorithm to sample from the posterior (A.10).

A.2.2 Imputation of Missing Type Indicators

The full conditional distribution of $\boldsymbol{\delta}_c$ is

$$\pi[\boldsymbol{\delta}_c \mid \cdot] \propto \prod_{d=0}^{1}\left\{\pi[\mathbf{x}_c^d \mid \cdot] \prod_{x\in\mathbf{x}_c^d} \pi(\delta_x = d)\right\} \propto \prod_{d=0}^{1}\prod_{x\in\mathbf{x}_c^d} \lambda_{1c}^d(x;\cdot)\pi(\delta_x = d).$$

Furthermore, the full conditional probability mass function of δ_x for each $x \in \mathbf{x}_c$ is $\pi[\delta_x = d \mid \cdot] \propto \lambda_{1c}^d(x;\cdot)\pi(\delta_x = d)$ for $d = 0, 1$. We set $\pi(\delta_x = d) = 0.5$, thus,

$$\pi[\delta_x = d \mid \cdot] \;=\; \frac{\lambda_{1c}^d(x;\cdot)}{\lambda_{1c}(x;\cdot)}, \qquad \text{for } d = 0, 1. \qquad \text{(A.11)}$$

A.2.3 Spatial Birth-and-Death Processes

The spatial birth-and-death process (Preston 1975, Møller and Waagepetersen 2004), a continuous time Markov process whose transitions are either births or deaths, can be used to simulate spatial point processes.

General Procedure

Suppose we wish to construct a spatial birth-and-death process to simulate a latent point process $\mathbf{A}$ from its posterior $\pi(\mathbf{a} \mid \mathbf{u})$, where $\mathbf{u}$ is data. Preston (1975) showed that if the birth and death rates satisfy the detailed balance equation

$$\pi(\mathbf{a} \mid \mathbf{u})b(\mathbf{a}, \xi) = \pi(\mathbf{a} \cup \{\xi\} \mid \mathbf{u})d(\mathbf{a} \cup \{\xi\}, \{\xi\}), \tag{A.12}$$

then the chain is time reversible and that the spatial birth-and-death process has a unique equilibrium distribution $\pi(\mathbf{a} \mid \mathbf{u})$ to which it converges in distribution from any initial state. In (A.12) $b(\mathbf{a}, \xi)$ is the birth rate for adding a new point ξ to the current configuration, $\mathbf{a}$, of the point process $\mathbf{A}$, and $d(\mathbf{a}, \xi)$ denotes the death rate for removing a point ξ from $\mathbf{a}$. We adopt the birth rate suggested by van Lieshout and Baddeley (2001) using a mixture intensity, i.e. $b(\mathbf{a}, \xi) = \sum_{u \in \mathbf{u}} h(\xi; u)$, where $h(\xi; \cdot)$ is an intensity function. To satisfy the detailed balance equation, the death rate for removing ξ from $\mathbf{a}$ is $d(\mathbf{a}, \xi) = [\pi(\mathbf{a}/\{\xi\} \mid \mathbf{u})/\pi(\mathbf{a} \mid \mathbf{u})] \sum_{u \in \mathbf{u}} h(\xi; u)$. Let $B(\mathbf{a}) = \int b(\mathbf{a}, \xi) d\xi$ denote the total birth rate. Note that $B(\mathbf{a})$ does not depend on $\mathbf{a}$. Write $B = B(\mathbf{a})$. Let $D(\mathbf{a}) = \sum_{a \in \mathbf{a}} d(\mathbf{a}; a)$ denote the total death rate. Given current state $\mathbf{a}$, after an exponentially distributed time with rate $B + D(\mathbf{a})$, a birth is proposed with probability $B/\{B + D(\mathbf{a})\}$ by sampling ξ from the mixture density $\sum_{u \in \mathbf{u}} h(\xi; u)/B$. A death is proposed with probability $D(\mathbf{a})/\{B + D(\mathbf{a})\}$ and the point $a \in \mathbf{a}$ is removed with probability $d(\mathbf{a}; a)/D(\mathbf{a})$.

Details for Simulating the Posterior Study Activation Center Processes

From (A.9), for $c = 1, \ldots, C$, the full conditional posterior distribution of a marked study activation center process for study c is

$$\pi[(\mathbf{y}, \mathbf{\Psi})_c \mid \cdot] \propto \pi[\mathbf{x}_c^1 \mid \eta_c, (\mathbf{y}, \mathbf{\Psi})_c] \times \pi[(\mathbf{y}, \mathbf{\Psi})_c \mid \rho_{2c}, (\mathbf{z}, \mathbf{\Sigma})],$$

for which we construct a spatial birth-and-death process using the birth rate

$$b_c[(\mathbf{y}, \mathbf{\Psi})_c, (y, \mathbf{\Psi}_y)] = \sum_{x \in \mathbf{x}_c^1} \eta_c \phi_3(y; x, \mathbf{\Psi}_y) \pi(\mathbf{\Psi}_y), \tag{A.13}$$

where $\phi_3(\xi; \mu, \mathbf{\Sigma})$ denotes the 3D normal density at ξ with mean μ and covariance matrix $\mathbf{\Sigma}$. The death rate for removing $(y, \mathbf{\Psi}_y) \in (\mathbf{y}, \mathbf{\Psi})_c$ is then

$$\begin{aligned}
& d_c[(\mathbf{y}, \mathbf{\Psi})_c, (y, \mathbf{\Psi}_y)] \\
&\quad = \frac{\pi[\mathbf{x}_c^1 \mid \eta_c, (\mathbf{y}, \mathbf{\Psi})_c / \{(y, \mathbf{\Psi}_y)\}] \cdot \pi[(\mathbf{y}, \mathbf{\Psi})_c / \{(y, \mathbf{\Psi}_y)\} \mid \rho_{2c}, (\mathbf{z}, \mathbf{\Sigma})]}{\pi[\mathbf{x}_c^1 \mid \eta_c, (\mathbf{y}, \mathbf{\Psi})_c] \cdot \pi[(\mathbf{y}, \mathbf{\Psi})_c \mid \rho_{2c}, (\mathbf{z}, \mathbf{\Sigma})]} \\
&\qquad \times \sum_{x \in \mathbf{x}_c} \eta_c \phi_3(y; x, \mathbf{\Psi}_y) \pi(\mathbf{\Psi}_y) \\
&\quad = \frac{\exp\{\eta_c \Phi_3(\cdot; y, \mathbf{\Psi}_y)\}}{\prod_{x \in \mathbf{x}_c^1} \left[1 + \frac{\eta_c \phi_3(x; y, \mathbf{\Psi}_y)}{\lambda_{1c}^1(x; \cdot) - \eta_c \phi_3(x; y, \mathbf{\Psi}_y)}\right]} \cdot \frac{\eta_c \sum_{x \in \mathbf{x}_c} \phi_3(y; x, \mathbf{\Psi}_y)}{\lambda_{2c}(y; \cdot)}, \qquad \text{(A.14)}
\end{aligned}$$

where $\Phi_3(\cdot \mid \mu, \mathbf{\Sigma}) = \int_{\mathcal{B}} \phi_3(\xi \mid \mu, \mathbf{\Sigma}) d\xi$, is 3D normal probability over the brain $\mathcal{B}$ with mean μ and covariance matrix $\mathbf{\Sigma}$. Using (A.13), the total birth rate is

$$\begin{aligned}
B_c &= \int_{\mathcal{B}} \int_{\mathcal{M}} \sum_{x \in \mathbf{x}_c^1} \eta_c \phi_3(y; x, \mathbf{\Psi}_y) \pi(\mathbf{\Psi}_y) d\mathbf{\Psi}_y dy \\
&= \eta_c \sum_{x \in \mathbf{x}_c^1} \mathbf{T}_3(\cdot; x, \boldsymbol{S}/(d-2), d), \qquad \text{(A.15)}
\end{aligned}$$

where $\mathcal{M}$ is the space of 3×3 symmetric positive definite matrices. $\mathbf{T}_3(\cdot; \mu, \boldsymbol{S}, d)$ is student-t probability over the brain $\mathcal{B}$ with mean μ, scale matrix $\boldsymbol{S}$ and d degrees of freedom. The total death rate is given by

$$D_c[(\mathbf{y}, \mathbf{\Psi})_c] = \sum_{(y, \mathbf{\Psi}_y) \in (\mathbf{y}, \mathbf{\Psi})_c} d_c[(\mathbf{y}, \mathbf{\Psi})_c, (y, \mathbf{\Psi}_y)]. \qquad \text{(A.16)}$$

For a birth, we draw a new point $(y, \mathbf{\Psi}_y)$ from

(A.17)

$$\frac{b_c[(\mathbf{y}, \mathbf{\Psi})_c, (y, \mathbf{\Psi}_y)]}{B_c} = \sum_{x \in \mathbf{x}_c^1} \left\{ \frac{\phi_3(y; x, \mathbf{\Psi}_y) \pi(\mathbf{\Psi}_y)}{\mathbf{T}_3(\cdot; x, \boldsymbol{S}/(d-2), d)} \cdot \frac{\mathbf{T}_3(\cdot; x, \boldsymbol{S}/(d-2), d)}{\sum_{x' \in \mathbf{x}_c^1} \mathbf{T}_3(\cdot; x', \boldsymbol{S}/(d-2), d)} \right\},$$

To draw from this mixture distribution, first draw

$$\mathbf{\Psi}_y \sim W^{-1}(\boldsymbol{S}, d), \qquad \text{(A.18)}$$

then draw y from the following mixture distribution,

$$[y \mid \mathbf{\Psi}_y] \sim \sum_{x \in \mathbf{x}_c^1} v_x N_{\mathcal{B}}(x, \mathbf{\Psi}_y), \qquad \text{(A.19)}$$

where $N_{\mathcal{B}}(\mu, \mathbf{\Sigma})$ denotes the 3D normal distribution truncated to $\mathcal{B}$ with mean μ and covariance matrix $\mathbf{\Sigma}$, and the weight $v_x = \mathbf{T}_3(\cdot; x, \boldsymbol{S}/(d-2), d)/B_c$.

The simulation time for this spatial birth-and-death process at each iteration of the hybrid algorithm is set to $1/B_c$, and the number of points in $\mathbf{y}_c$ is initially set to zero.

Details for Simulating the Posterior Population Center Process

Note that (A.6) implies that the intensity functions for $\mathbf{X}_0^c$, $\mathbf{Y}_c$ and $\mathbf{Y}$, say, λ_{1c}^0, λ_{2c} and λ have the following relationships: $\lambda_{1c}^0(y; \cdot) = \rho_{1c}\lambda(y; \cdot)$ and $\lambda_{2c}(y; \cdot) = \rho_{2c}\lambda(y; \cdot)$. Thus,

$$\begin{aligned}
&\prod_{c=1}^{C} \left\{ \pi[\mathbf{x}_c^0 \mid \rho_{1c}, (\mathbf{z}, \mathbf{\Sigma})] \cdot \pi[\mathbf{y}_c \mid \rho_{2c}, (\mathbf{z}, \mathbf{\Sigma})] \right\} \\
&\quad \propto \prod_{c=1}^{C} \left\{ \exp\left\{ -\int_{\mathcal{B}} \lambda_{1c}^0(x; \cdot) + \lambda_{2c}(x; \cdot) dx \right\} \prod_{x \in \mathbf{x}_c^0} \lambda_{1c}^0(x; \cdot) \prod_{y \in \mathbf{y}_c} \lambda_{2c}(y; \cdot) \right\} \\
&\quad \propto \pi[\mathbf{y} \mid (\mathbf{z}, \mathbf{\Sigma})] \cdot \prod_{c=1}^{C} \left\{ \rho_{1c}^{\mathbf{n}(\mathbf{x}_c^0)} \rho_{2c}^{\mathbf{n}(\mathbf{y}_c)} \right\}.
\end{aligned}$$

This implies that (A.10), the full joint posterior, is proportional to

$$\begin{aligned}
&\prod_{c=1}^{C} \left\{ \pi\left[\mathbf{x}_c^1 \mid \eta_c, (\mathbf{y}_c, \mathbf{\Psi}_c)\right] \pi(\eta_c) \rho_{1c}^{\mathbf{n}(\mathbf{x}_c^0)} \rho_{2c}^{\mathbf{n}(\mathbf{y}_c)} \right\} \prod_{y \in \mathbf{y}} \pi(\mathbf{\Psi}_y) \times \\
&\pi(\boldsymbol{\rho}) \pi\left[\mathbf{y} \mid (\mathbf{z}, \mathbf{\Sigma})\right] \pi\left[\mathbf{z} \mid \beta\right] \pi(\beta) \pi(\boldsymbol{T}) \prod_{z \in \mathbf{z}} \pi(\mathbf{\Sigma} \mid \boldsymbol{T}).
\end{aligned} \tag{A.20}$$

Thus, the full conditional posterior distribution of the population center process is

$$\pi[(\mathbf{z}, \mathbf{\Sigma}) \mid \cdot] \quad \propto \quad \pi[\mathbf{y} \mid (\mathbf{z}, \mathbf{\Sigma})] \pi[\mathbf{z} \mid \beta] \prod_{z \in \mathbf{z}} \pi(\mathbf{\Sigma}_z \mid \boldsymbol{T}),$$

for which we construct a spatial birth-and-death process using the birth rate,

$$b[(\mathbf{z}, \mathbf{\Sigma}), (z, \mathbf{\Sigma}_z)] = \beta \pi(\mathbf{\Sigma}_z \mid \boldsymbol{T}) \left[1 + \frac{\theta}{\epsilon} \sum_{y \in \mathbf{y}} \phi_3(y; z, \mathbf{\Sigma}_z) \right], \tag{A.21}$$

and the death rate, for removing a point $(z, \mathbf{\Sigma}_z) \in (\mathbf{z}, \mathbf{\Sigma})$, is

$$d[(\mathbf{z}, \mathbf{\Sigma}), (z, \mathbf{\Sigma}_z)] = \frac{\exp\{\theta \Phi_3(\cdot; z, \mathbf{\Sigma}_z)\}}{\prod_{y \in \mathbf{y}} \left[1 + \frac{\theta \phi_3(y; z, \mathbf{\Sigma}_z)}{\lambda(y; \cdot) - \theta \phi_3(y; z, \mathbf{\Sigma}_z)}\right]} \left[1 + \frac{\theta}{\epsilon} \sum_{y \in \mathbf{y}} \phi_3(y; z, \mathbf{\Sigma}_z)\right]. \tag{A.22}$$

The total birth rate is

$$B = \int_{\mathcal{B}} \int_{\mathcal{M}} b[(\mathbf{z}, \mathbf{\Sigma}), (z, \mathbf{\Sigma}_z)] d\mathbf{\Sigma}_z dz = \beta \left[|\mathcal{B}| + \frac{\theta}{\epsilon} \sum_{y \in \mathbf{y}} \mathbf{T}_3(\cdot; y, \boldsymbol{T}/(\nu - 2), \nu - 2)\right] \tag{A.23}$$

and the total death rate is

$$D(\mathbf{z}, \mathbf{\Sigma}) = \sum_{z \in \mathbf{z}} d[(\mathbf{z}, \mathbf{\Sigma}), (z, \mathbf{\Sigma}_z)] \tag{A.24}$$

For a birth, we draw a new point $(z, \mathbf{\Sigma}_z)$ from the mixture density

$$\begin{aligned} &\frac{b[(\mathbf{z}, \mathbf{\Sigma}), (z, \mathbf{\Sigma}_z)]}{B} \\ &= \sum_{y \in \mathbf{y}} \left\{ \frac{\beta |\mathcal{B}| \pi(\mathbf{\Sigma}_z)}{B} + \frac{\phi_3(y; z, \mathbf{\Sigma}_z) \pi(\mathbf{\Sigma}_z)}{\mathbf{T}_3(\cdot; y, \frac{\boldsymbol{T}}{\nu - 2}, \nu - 2)} \frac{\beta \mathbf{T}_3(\cdot; y, \frac{\boldsymbol{T}}{\nu - 2}, \nu - 2)}{B} \right\}. \end{aligned} \tag{A.25}$$

This implies that we first draw

$$\mathbf{\Sigma}_z \sim W^{-1}(\boldsymbol{S}, \nu), \tag{A.26}$$

then draw z from the following mixture distribution,

$$[z \mid \mathbf{\Sigma}_z] \sim w_{\emptyset} U_{\mathcal{B}} + \sum_{y \in \mathbf{y}} w_y N_{\mathcal{B}}(y, \mathbf{\Sigma}_z), \tag{A.27}$$

where $U_{\mathcal{B}}$ denotes the uniform distribution over the brain $\mathcal{B}$, and the weight $w_y = \beta |\mathcal{B}| / B$ when $y = \emptyset$ and $w_y = \beta B^{-1} \mathbf{T}_3(\cdot; y, \boldsymbol{T}/(\nu - 2), \nu - 2)$ if $y \in \mathbf{y}$.

The total simulation time for this spatial birth-and-death process at each iteration of the hybrid algorithm is set to $1/B$ and number of points in the population center process is set to zero as the initial state.

A.2.4 Standard MCMC Updates

Upon exiting the spatial birth-and-death processes, conditional on the number of points in $\mathbf{Y}_c$ and $\mathbf{Z}$, we use a standard MCMC algorithm to update the study activation center processes parameters and population center process parameters, i.e. $(\mathbf{y}, \boldsymbol{\Psi})_c$ and $(\mathbf{z}, \boldsymbol{\Sigma})$, as well as $\boldsymbol{\eta}$, $\boldsymbol{\rho}$, β and $\boldsymbol{T}$.

Update $(\mathbf{y}, \boldsymbol{\Psi})_c$ given $N_{\mathbf{Y}_c}(\mathcal{B}) = n_c$

Define a latent indicator variable $\alpha_x \in \mathbf{y}_c$ with prior $\pi(\alpha_x = y) = 1/n_c$ for each $x \in \mathbf{X}_c^1$ and $y \in \mathbf{y}_c$, such that $\mathbf{X}_{cy}^1 = \{x \in \mathbf{X}_c^1; \alpha_x = y\}$. Recall that $\mathbf{X}_{cy}^1$ is a process defined in Section 2.1 with intensity function $\eta_c \phi_3(x; y, \boldsymbol{\Psi}_y)$. Write $\boldsymbol{\alpha}_c = \{\alpha_x, x \in \mathbf{X}_c^1\}$. Note that $\pi[\mathbf{x}_c^1, \boldsymbol{\alpha}_c \mid \lambda_{1c}^1] \propto \prod_{y \in \mathbf{y}} \pi[\mathbf{x}_{cy} \mid \eta_c, y, \boldsymbol{\Psi}_y]$. Then the full conditional distribution of $(\mathbf{y}, \boldsymbol{\Psi})_c$ and $\boldsymbol{\alpha}_c$ is $\pi[(\mathbf{y}, \boldsymbol{\Psi})_c, \boldsymbol{\alpha}_c \mid \cdot] \propto \prod_{y \in \mathbf{y}_c} \Big\{\pi[\mathbf{x}_{cy} \mid \eta_c, \boldsymbol{\Psi}_y] \pi(\boldsymbol{\Psi}_y)\Big\} \pi[\mathbf{y}_c \mid \rho_{2c}, (\mathbf{z}, \boldsymbol{\Sigma})]$. This further implies the full conditional of $y \in \mathbf{y}_c$ is

$$\pi[y \mid \cdot] \propto \exp\{-\eta_c \Phi_3(\cdot; y, \boldsymbol{\Psi}_y)\} \lambda_{2c}(y; \cdot)\, \phi_3 \left(y; \frac{\sum\limits_{x \in \mathbf{x}_c^1} x I(\alpha_x = y)}{\sum\limits_{x \in \mathbf{x}_c^1} I(\alpha_x = y)}, \frac{\boldsymbol{\Psi}_y}{\sum\limits_{x \in \mathbf{x}_c^1} I(\alpha_x = y)} \right).$$

Thus to update $y \in \mathbf{y}_c$, draw

$$y^* \sim N_{\mathcal{B}} \left[\frac{\sum_{x \in \mathbf{x}_c^1} x I(\alpha_x = y)}{\sum_{x \in \mathbf{x}_c^1} I(\alpha_x = y)}, \frac{\boldsymbol{\Psi}_y}{\sum_{x \in \mathbf{x}_c^1} I(\alpha_x = y)} \right], \tag{A.28}$$

and accept with probability

$$\min \left\{ 1, \exp\{\eta_c [\Phi_3(\cdot; y, \boldsymbol{\Psi}_y) - \Phi_3(\cdot; y^*, \boldsymbol{\Psi}_y)]\} \frac{\lambda_2(y^*; \cdot)}{\lambda_2(y; \cdot)} \right\}. \tag{A.29}$$

The full conditional of $\boldsymbol{\Psi}_y$ is

$$\begin{aligned} \pi[\boldsymbol{\Psi}_y \mid \cdot] \;\; \propto \;\; & \exp\{-\eta_c \Phi_3(\cdot; y, \boldsymbol{\Psi}_y)\} \pi(\boldsymbol{\Psi}_y) \\ & \times \phi_3 \left(y; \frac{\sum\limits_{x \in \mathbf{x}_c^1} x I(\alpha_x = y)}{\sum\limits_{x \in \mathbf{x}_c^1} I(\alpha_x = y)}, \frac{\boldsymbol{\Psi}_y}{\sum\limits_{x \in \mathbf{x}_c^1} I(\alpha_x = y)} \right). \end{aligned}$$

Thus to update $\mathbf{\Psi}_y$, draw

$$\text{(A.30)}\quad \mathbf{\Psi}_y^* \sim W^{-1}\left[\boldsymbol{S} + \sum_{x\in\mathbf{x}_1^c}(x-y)(x-y)^{\mathrm{T}} I(\alpha_x = y), d + \sum_{x\in\mathbf{x}_c^1} I(\alpha_x = y)\right],$$

and accept with probability

$$\text{(A.31)}\qquad \min\left\{1, \exp\{-\eta_c[\Phi_3(\cdot; y, \mathbf{\Psi}_y^*) - \Phi_3(\cdot; y, \mathbf{\Psi}_y)]\}\right\}.$$

To update α_x, we have

$$\text{(A.32)}\qquad \pi[\alpha_x = y \mid \cdot] \;=\; \frac{\phi_3(x; y, \mathbf{\Psi}_y)}{\sum_{y'\in\mathbf{y}_c}\phi_3(x; y', \mathbf{\Psi}_{y'})} \quad \forall\, y \in \mathbf{y}_c.$$

Update $(\mathbf{z}, \mathbf{\Sigma})$ given $N_{\mathbf{Z}}(\mathcal{B}) = m$

Define a latent indicator variable $\gamma_y \in \mathbf{z}\cup\{\emptyset\}$ with prior $\pi(\gamma_y = z) = 1/(m+1)$ for each $z \in \mathbf{z}\cup\{\emptyset\}$ and $y \in \mathbf{Y}$ such that $\mathbf{Y}_{cz} = \{y \in \mathbf{Y}; \gamma_y = z\}$. From Section 2.1, $\mathbf{Y}_{c\emptyset}$ is a homogeneous Poisson process with constant intensity ϵ and $\mathbf{Y}_{cz}$ for $z \in \mathbf{Z}$ is a point process with intensity $\theta\phi_3(y; z, \mathbf{\Sigma}_z)$. Write $\boldsymbol{\gamma} = \{\gamma_y, y \in \mathbf{Y}\}$. Note that $\pi[(\mathbf{y}, \boldsymbol{\gamma}) \mid \cdot] \propto \pi[\mathbf{y}_{c\emptyset}]\prod_{z\in\mathbf{z}}\pi[\mathbf{y}_{cz} \mid z, \mathbf{\Sigma}_z]$. Then the joint posterior distribution of $(\mathbf{z}, \mathbf{\Sigma})$ and $\boldsymbol{\gamma}$ given all other parameters is

$$\pi[(\mathbf{z}, \mathbf{\Sigma}), \boldsymbol{\gamma} \mid \cdot] \propto \prod_{z\in\mathbf{z}}\left\{\pi[\mathbf{y}_{cz} \mid z, \mathbf{\Sigma}_z]\pi[\mathbf{\Sigma}_z]\right\}\pi[\mathbf{z} \mid \beta].$$

Thus the full conditional of $z \in \mathbf{z}$ is

$$\pi[z \mid \cdot] \propto \exp\{-\theta\Phi_3(\cdot; z, \mathbf{\Sigma}_z)\}\phi_3\left(z; \frac{\sum_{y\in\mathbf{y}} yI(\gamma_y = z)}{\sum_{y\in\mathbf{y}} I(\gamma_y = z)}, \frac{\mathbf{\Sigma}_z}{\sum_{y\in\mathbf{y}} I(\gamma_y = z)}\right).$$

Thus to update $z \in \mathbf{z}$, draw

$$\text{(A.33)}\qquad z^* \sim N_{\mathcal{B}}\left[\frac{\sum_{y\in\mathbf{y}} yI(\gamma_y = z)}{\sum_{y\in\mathbf{y}} I(\gamma_y = z)}, \frac{\mathbf{\Sigma}_z}{\sum_{y\in\mathbf{y}} I(\gamma_y = z)}\right],$$

and accept with probability

$$\text{(A.34)}\qquad \min\{1, \exp\{\theta[\Phi_3(\cdot, z, \mathbf{\Sigma}_z) - \Phi_3(\cdot, z^*, \mathbf{\Sigma}_z)]\}\}.$$

The full conditional of $\mathbf{\Sigma}_z$ is

$$\begin{aligned}\pi[\mathbf{\Sigma}_z \mid \cdot] &\propto \exp\{-\theta\Phi_3(\cdot; z, \mathbf{\Sigma}_z)\}\pi(\mathbf{\Sigma}_z \mid \boldsymbol{T}) \\ &\qquad\times\phi_3\left(z; \frac{\sum_{y\in\mathbf{y}} yI(\gamma_y = z)}{\sum_{y\in\mathbf{y}} I(\gamma_y = z)}, \frac{\mathbf{\Sigma}_z}{\sum_{y\in\mathbf{y}} I(\gamma_y = z)}\right).\end{aligned}$$

Thus update $\mathbf{\Sigma}_z$ by first drawing

$$\text{(A.35) } \mathbf{\Sigma}_z^* \sim W^{-1}\left[\boldsymbol{T} + \sum_{y\in\mathbf{y}}(y-z)(y-z)^{\mathrm{T}}I(\gamma_y = z), \nu + \sum_{y\in\mathbf{y}} I(\gamma_y = z)\right],$$

and accept with probability

$$\text{(A.36)} \qquad \min\{1, \exp\{\theta[\Phi_3(\cdot, z, \mathbf{\Sigma}_z) - \Phi_3(\cdot, z, \mathbf{\Sigma}_z^*)]\}\}.$$

To update each γ_y for $y \in \mathbf{Y}$, we have

$$\text{(A.37)} \qquad \pi[\gamma_y = \emptyset \mid \cdot] = \frac{\epsilon}{\epsilon + \theta\sum_{z'\in\mathbf{z}}\phi_3(y; z', \mathbf{\Sigma}_{z'})}$$

$$\text{(A.38)} \qquad \pi[\gamma_y = z \mid \cdot] = \frac{\theta\phi_3(y; z, \mathbf{\Sigma}_z)}{\epsilon + \theta\sum_{z'\in\mathbf{z}}\phi_3(y; z', \mathbf{\Sigma}_{z'})} \quad \forall\, z \in \mathbf{z},$$

Update $\boldsymbol{\eta}$, $\boldsymbol{\rho}$, β and $\boldsymbol{T}$

The full conditional for η_c for $c = 1, 2, \ldots, C$ is

$$\begin{aligned}&\text{(A.39)} & \pi[\eta_c \mid \cdot] &\propto \pi[\mathbf{x}_c^1 \mid \eta_c, (\mathbf{y}, \mathbf{\Psi})_c]\pi(\eta_c) \\ &\text{(A.40)} & &\propto \exp\left\{-\eta_c\left[\sum_{y\in\mathbf{y}_c}\Phi_3(\cdot; y, \mathbf{\Psi}_y) + b_\eta\right]\right\}\eta_c^{\mathbf{n}(\mathbf{x}_c^1)+a_\eta},\end{aligned}$$

Thus to update η_c draw

$$\text{(A.41)} \qquad \eta_c \sim G\left[\sum_{y\in\mathbf{y}_c}\Phi_3(\cdot; y, \mathbf{\Psi}_y) + b_\eta, \mathbf{n}(\mathbf{x}_c^1) + a_\eta\right],$$

According to (A.20), the full conditional for $\boldsymbol{\rho}$ is

$$\text{(A.42)} \quad \pi[\boldsymbol{\rho} \mid \cdot] \propto \prod_{c=1}^{C}\left\{\rho_{1c}^{\mathbf{n}(\mathbf{x}_c^0)}\rho_{2c}^{\mathbf{n}(\mathbf{y}_c)}\right\}\pi(\boldsymbol{\rho}) \propto \prod_{c=1}^{C}\left\{\rho_{1c}^{\mathbf{n}(\mathbf{x}_c^0)+\alpha_{1c}}\rho_{2c}^{\mathbf{n}(\mathbf{y}_c)+\alpha_{2c}}\right\},$$

Thus update $\boldsymbol{\rho}$ by drawing

$$\boldsymbol{\rho} \sim D[\mathbf{n}(\mathbf{x}_1^0) + \alpha_{11}, \ldots, \mathbf{n}(\mathbf{x}_C^0) + \alpha_{1C}, \mathbf{n}(\mathbf{y}_1) + \alpha_{21}, \ldots, \mathbf{n}(\mathbf{y}_C) + \alpha_{2C}]. \tag{A.43}$$

The full conditional for β is $\pi[\beta \mid \cdot] \propto \exp\{-\beta(|\mathcal{B}| + b_\beta)\}\, \beta^{\mathbf{n}(\mathbf{z})+a_\beta}$. Thus draw

$$\beta \sim G[\mathbf{n}(\mathbf{z}) + \alpha_\beta, |\mathcal{B}| + b_\beta]. \tag{A.44}$$

The full conditional of $\boldsymbol{T}$ is $\pi[\boldsymbol{T} \mid \cdot] \propto \pi(\boldsymbol{T}) \prod_{z\in\mathbf{z}} \pi(\boldsymbol{\Sigma}_z \mid \boldsymbol{T})$. Therefore, draw

$$\boldsymbol{T} \sim W^{-1}\left[(\boldsymbol{T}_0^{-1} + \sum_{z\in\mathbf{z}} \boldsymbol{\Sigma}_z^{-1}), \nu_0 + \nu\mathbf{n}(\mathbf{z})\right]. \tag{A.45}$$

A.2.5 Normal and T Probability Computation

The total birth rate and death rate in the spatial birth-and-death process, as well as many standard MCMC updates involve the evaluation of 3D normal and student-t probabilities over the brain. It is difficult to directly evaluate these probabilities over arbitrary regions such as the brain. Thus, we resort to Monte Carlo simulation of these probabilities. Since more than half of the normal and t probabilities in our motivating example are close to 1 we consider the following approximation. If the 99% credible ellipsoid of the target distribution lies completely within the brain we set the probability to 0.995, otherwise, we estimate it via Monte Carlo simulation. We set the Monte Carlo sample size $n = 500$ using the estimated standard error, $\sqrt{p(1-p)/n}$, where p is the true probability. This results in a maximum absolute value of the Monte Carlo error (twice the standard error) of 0.05 when $p = 0.5$.

A.2.6 Posterior Intensity Estimation

Let $(\mathbf{z}^{(k)}, \boldsymbol{\Sigma}^{(k)})$ be the posterior draw of the population center process at the kth iteration after burn-in, for $k = 1, \ldots, K$. To obtain its intensity function, we combine all of the draws of locations, i.e. $\cup_{k=1}^K \mathbf{z}^{(k)}$, which are then smoothed with a mixture of Dirichlet process priors model (a non-parametric Bayesian model used for density smoothing/estimation, Escobar and West (1995)). We rescale the density by multiplying the posterior mean number of population centers, i.e. $\sum_{k=1}^{K} \mathbf{n}(\mathbf{z}^{(k)})/K$ to estimate the intensity.

Also, using $(\mathbf{z}^{(k)}, \mathbf{\Sigma}^{(k)})$, we estimate the activation center process intensity function by $K^{-1} \sum_{k=1}^{K} \lambda(y; \mathbf{z}^{(k)}, \mathbf{\Sigma}^{(k)})$, where $\lambda(y; \cdot)$ is defined by equation (9) in Section 2.2.

Let $\boldsymbol{\rho}^{(k)}$ be the posterior draw at the kth iteration after burn-in. To obtain a posterior predictive intensity $\tilde{\lambda}$ for a new study, given $\boldsymbol{\rho}^{(k)}$ and $(\mathbf{z}^{(k)}, \mathbf{\Sigma}^{(k)})$, we randomly pick $c^{(k)} \in \{1, \ldots, C\}$, then simulate a Poisson point process $\mathbf{y}_0^{(kl)}$ with intensity $\rho_{1c^{(k)}}^{(k)} \lambda(y; \mathbf{z}^{(k)}, \mathbf{\Sigma}^{(k)})$, for $l = 1, 2, \ldots, L$. For the kth iteration, $\tilde{\lambda}^{(k)}$ with a voxel v is estimated by

$$\tilde{\lambda}^{(k)}(y; \cdot) = \frac{\sum_{l=1}^{L} \sum_{y_0 \in \mathbf{y}^{(kl)}} I[y_0 \in v]}{L|v|}, \text{ for } y \in v,$$

Then $\tilde{\lambda}(y;)$ is estimated by $\sum_{k=1}^{K} \tilde{\lambda}^{(k)}(y; \cdot)/K$.

A.2.7 Credible Ellipsoid Computation

Conditional on the event that there is exactly one population center, $z \in \mathbf{Z}$, in the region of interest, the posterior distribution of z can be approximated by a normal distribution $N(\mu_z, \mathbf{\Lambda}_z)$, where $(\mu_z, \mathbf{\Lambda}_z)$ can be simulated via a mixture of Dirichlet process priors model based on the posterior sample of $\mathbf{Z}$. The 95% credible ellipsoid for the population centers is $\mathrm{CR}_z^p = \{x : (x - \mu_z)^{\mathrm{T}} \mathbf{\Lambda}_z^{-1} (x - \mu_z) \le \chi^2_{0.95,3}\}$, where $\chi^2_{0.95,3}$ is the 0.95 quantile of χ^2 distribution with 3 degrees of freedom.

Conditional on the same event, the posterior distribution of an activation center $y \in (\mathbf{Y}_z = \cup_{c=1}^{C} \mathbf{Y}_{cz})$ that is associated with the population center z located in the region of interest, is a normal distribution with mean z and covariance matrix $\mathbf{\Sigma}_z$, which can be estimated using the posterior draws of the population center process.

Thus, the 95% credible ellipsoid for the activation centers is $\mathrm{CR}_z^a = \{x : (x - z)^{\mathrm{T}} \boldsymbol{\Sigma}_z^{-1} (x - z) \leq \chi^2_{0.95,3}\}$.

A.2.8 Pseudo Code

Starting with an initial state of all the parameters, repeat the following steps for a pre-specified total number of iterations:

```
1. Update $\boldsymbol{\delta}$ according to (A.11).
2. For $c = 1, \ldots, C$, run the spatial birth-and-death process for study center processes.
    2.1. Compute $B_c$ according to (A.15); set $\tau_c = 1/B_c$; set $t = 0$; set $(\mathbf{y}, \boldsymbol{\Psi})_c = (\{\emptyset\}, \{\emptyset\})$;
    2.2. Compute $d_c[\cdot, (y, \boldsymbol{\Psi}_y)]$ for all $(y, \boldsymbol{\Psi}_y) \in (\mathbf{y}, \boldsymbol{\Psi})_c$ and $D_c$ according to (A.14) and (A.16).
    2.3. Draw $r \sim U[0, 1]$. If $r < B_c/(B_c + D_c)$, then draw $(y, \boldsymbol{\Psi}_y)$ according to (A.18) and (A.19), and set $(\mathbf{y}, \boldsymbol{\Psi})_c = (\mathbf{y} \cup \{y\}, \boldsymbol{\Psi} \cup \{\boldsymbol{\Psi}_y\})_c$, else select $(y, \boldsymbol{\Psi}_y)$ from $(\mathbf{y}, \boldsymbol{\Psi})_c$ with probability $d_c[\cdot, (y, \boldsymbol{\Psi}_y)]/D_c$ and set $(\mathbf{y}, \boldsymbol{\Psi})_c = (\mathbf{y} \backslash \{y\}, \boldsymbol{\Psi} \backslash \{\boldsymbol{\Psi}_y\})_c$.
    2.4. Draw a sojourn time $s$ from an exponential distribution with rate $B_c + D_c$; set $t = t + s$, if $t < \tau_c$, then go to 2.2.
3. Update each $(y, \boldsymbol{\Psi}_y) \in (\mathbf{y}, \boldsymbol{\Psi})_c$, for $c = 1, \ldots, C$ based on (A.28) and (A.30).
4. Update $\eta_c$, for $c = 1, \ldots, C$ according to (A.41).
5. Update $\boldsymbol{\rho}$ from (A.43).
6. Run the spatial birth-and-death process for the posterior population center process:
```

6.1. Compute B based on (A.23); set $\tau_c = 1/B$; set $t = 0$; set $(\mathbf{z}, \mathbf{\Sigma}) = (\{\emptyset\}, \{\emptyset\})$.

6.2. Compute $d[\cdot, (z, \mathbf{\Sigma}_z)]$ for all $(z, \mathbf{\Sigma}_z) \in (\mathbf{z}, \mathbf{\Sigma})$ and D based on (A.22) and (A.24).

6.3. Draw $r \sim U[0, 1]$. If $r < B/(B + D)$, then draw $(z, \mathbf{\Sigma}_z)$ according to (A.26) and (A.27), and set $(\mathbf{z}, \mathbf{\Sigma}) = (\mathbf{z} \cup \{z\}, \mathbf{\Sigma} \cup \{\mathbf{\Sigma}_z\})$, else select $(z, \mathbf{\Sigma}_z)$ from $(\mathbf{z}, \mathbf{\Sigma})$ with probability $d[\cdot, (z, \mathbf{\Sigma}_z)]/D$ and set $(\mathbf{z}, \mathbf{\Sigma}) = (\mathbf{z}\backslash\{z\}, \mathbf{\Sigma}\backslash\{\mathbf{\Sigma}_z\})$.

6.4. Draw a sojourn time s from an exponential distribution with rate $B + D$; set $t = t + s$; if $t < \tau$, then go to 6.2.

7. Update each $(z, \mathbf{\Sigma}_z) \in (\mathbf{z}, \mathbf{\Sigma})$ based on (A.33) and (A.35).

8. Update β from (A.44).

9. Update $\boldsymbol{T}$ from (A.45).

A.3 Simulation Studies

In this section, we conduct two simulation studies. In the first, we investigate the sensitivity of the prior parameter settings in our model for the emotion data. In the second, we study sensitivity of the model specification by simulating three data sets based on different models.

A.3.1 Sensitivity to Priors

Our primary interest is how the posterior inference (including the number, location and variability) of the population centers varies with different informative prior specifications. We keep the non-informative priors that $\boldsymbol{\rho} \sim D(0.5, \ldots, 0.5)$ and $\eta_c \sim G(0.001, 0.001)$ for any $c = 1, 2, \cdots, C$ and let $\beta|\mathcal{B}| \sim G(a_\beta, 0.001)$, which implies that $\mathrm{var}(\beta|\mathcal{B}|)/E(\beta|\mathcal{B}|) = 1000$, thus the prior for β is relatively vague. We consider nine scenarios where scenario 1 has the same prior set up as in Section 3 in the manuscript, other scenarios vary the settings of $E[N_{\mathbf{Z}}(\mathcal{B})]$, $E[N_{\mathbf{A}}(\mathcal{B})]/E[N_{\mathbf{Y}}(\mathcal{B})]$, $E[N_{\mathbf{U}_c}(\mathcal{B})]$, $E[\boldsymbol{\Sigma}_z]$ and $E[\boldsymbol{\Psi}_y]$ (see Table A.1 for a summary of the different scenarios), where $\mathbf{U}_c = \mathbf{Y}_c \cup \mathbf{X}_c^0$, is the activation centers in study c, $\mathbf{Y} = \cup_{c=1}^{C} \mathbf{U}_c$, represents the activation centers over all studies, and $\mathbf{A} = \cup_{c=1}^{C} \cup_{z \in \mathbf{z}} (\mathbf{X}_{cz} \cup \mathbf{Y}_{cz})$, denotes the activation centers that cluster at the population level.

Table A.1: Sensitivity Analysis Priors.

Scenario	${}^{a}E[N_{\mathbf{Z}}(\mathcal{B})]$	${}^{b}E[N_{\mathbf{U}_c}(\mathcal{B}))]$	${}^{c}\frac{E[N_{\mathbf{A}}(\mathcal{B})]}{E[N_{\mathbf{Y}}(\mathcal{B})]}$	$E[\boldsymbol{\Sigma}_z]$	$E[\boldsymbol{\Psi}_y]$	$\boldsymbol{T}_0$	${}^{d}\theta$	${}^{d}\epsilon\lvert\mathcal{B}\rvert$
1	30	5.0	0.80	$4\boldsymbol{I}$	$\boldsymbol{I}$	$0.8\boldsymbol{I}$	58.27	437.00
2	25	5.0	0.80	$4\boldsymbol{I}$	$\boldsymbol{I}$	$0.8\boldsymbol{I}$	69.92	437.00
3	35	5.0	0.80	$4\boldsymbol{I}$	$\boldsymbol{I}$	$0.8\boldsymbol{I}$	49.94	437.00
4	30	5.0	0.75	$4\boldsymbol{I}$	$\boldsymbol{I}$	$0.8\boldsymbol{I}$	54.62	546.25
5	30	5.0	0.85	$4\boldsymbol{I}$	$\boldsymbol{I}$	$0.8\boldsymbol{I}$	61.91	327.75
6	30	5.0	0.80	$8\boldsymbol{I}$	$\boldsymbol{I}$	$1.6\boldsymbol{I}$	58.27	437.00
7	30	5.5	0.80	$4\boldsymbol{I}$	$\boldsymbol{I}$	$0.8\boldsymbol{I}$	64.09	480.70
8	30	4.5	0.80	$4\boldsymbol{I}$	$\boldsymbol{I}$	$0.8\boldsymbol{I}$	52.44	393.30
9	30	5.0	0.80	$4\boldsymbol{I}$	$2\boldsymbol{I}$	$0.8\ \boldsymbol{I}$	58.27	437.00

[a] A priori expected number of population centers.
[b] A priori expected number of activation centers per study.
[c] A priori proportion of act. centers that cluster about a population center.
[d] Derived values from footnotes a, b and c.

We simulate the posterior distribution with 20,000 iterations after a burn-in of 2,000 iterations. Table A.2 shows descriptive statistics on the posterior distribution of the number of population centers after burn-in. As anticipated, scenarios 2 and 7 result in a decrease in the posterior mode of the number of the population centers. This is because scenario 2 has a smaller prior mean number of population centers while scenario 7 has a larger prior mean number of activation centers that clusters about population centers. However, the change is not dramatic. The posterior mode of the number of population centers for all nine scenarios is around 42, with a range from 36 to 49.

In Figure A.1 we compare sensitivity of the activation center posterior intensity, and in Figure A.2 we compare sensitivity of the population center posterior intensity.

Table A.2: The summary statistics on the posterior number of population centers

Scenario	Min	Max	Mean	s.d.	Mode
1	38	47	42.60	1.4	42
2	32	40	35.90	1.2	36
3	44	55	49.10	1.5	49
4	37	48	42.40	1.4	42
5	38	47	42.50	1.4	42
6	37	49	42.60	1.5	43
7	34	42	37.60	1.3	37
8	43	55	48.30	1.6	48
9	36	45	40.10	1.3	40

Figure A.3 compares the marginal credible ellipses for both population centers and activation centers. From these figures, we see that the posterior expected intensities and marginal credible ellipses are qualitatively quite similar. Table A.3 summarizes the estimated location (x, y, z), the Euclidean distance in locations between scenario 1 and other scenarios, the volume of the 95% credible ellipsoid and the volume of the 95% credible interval of population centers as well as activation centers conditional on the population centers located in the amygdalae. The estimated number of population centers and the estimated volume of the credible ellipsoids is somewhat sensitive to the choice of prior on β, θ and ϵ, but the posterior intensity and the estimated location of the population centers are stable to various prior specifications. The maximum distance in credible ellipsoid locations for population center between scenario 1 and other scenarios is around 1mm (0.5 voxel).

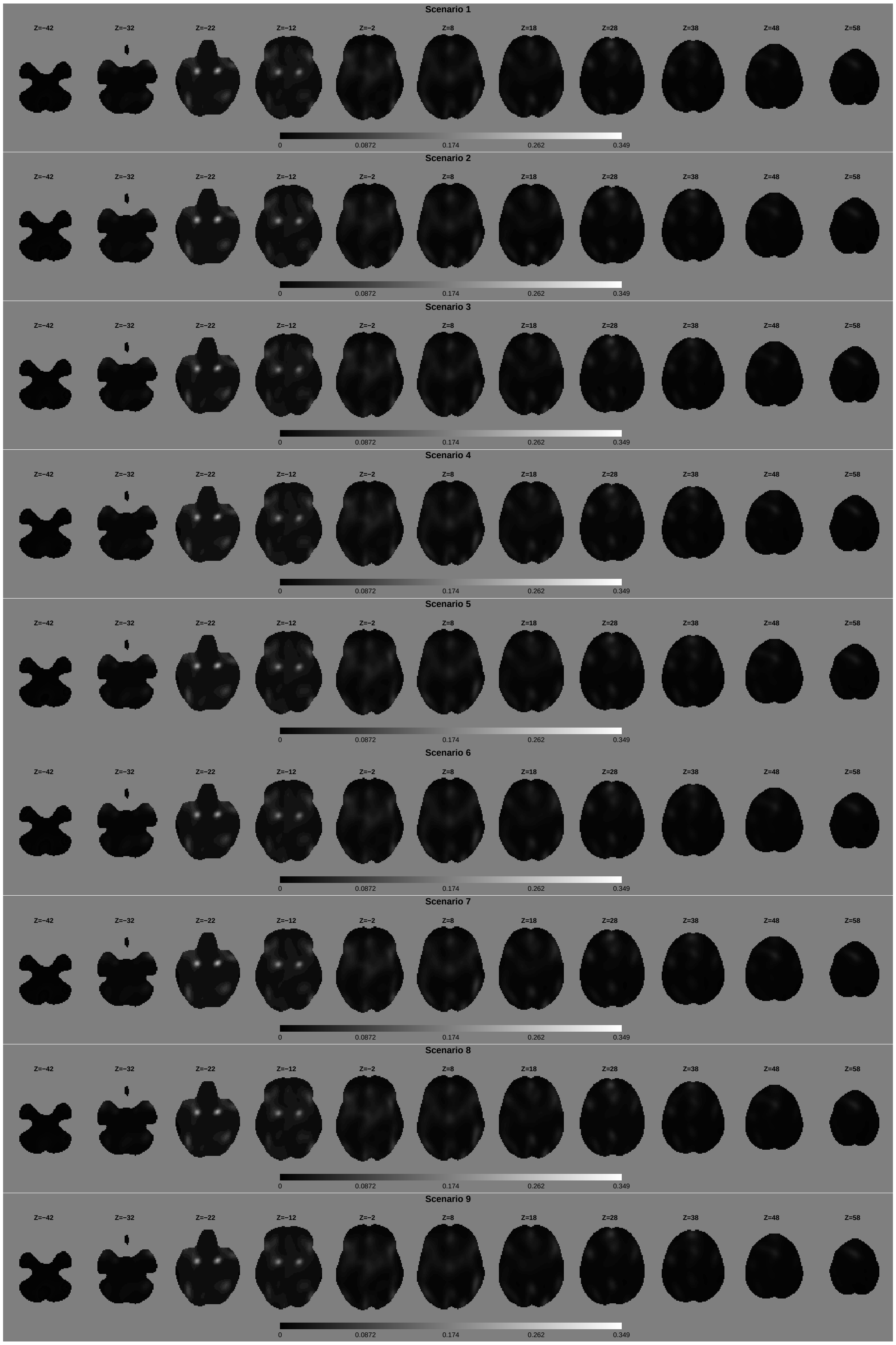


Figure A.1: Sensitivity analysis results: comparisons of the posterior activation center intensity on 11 slices of the brain from $Z = -42$mm to $Z = 58$mm.

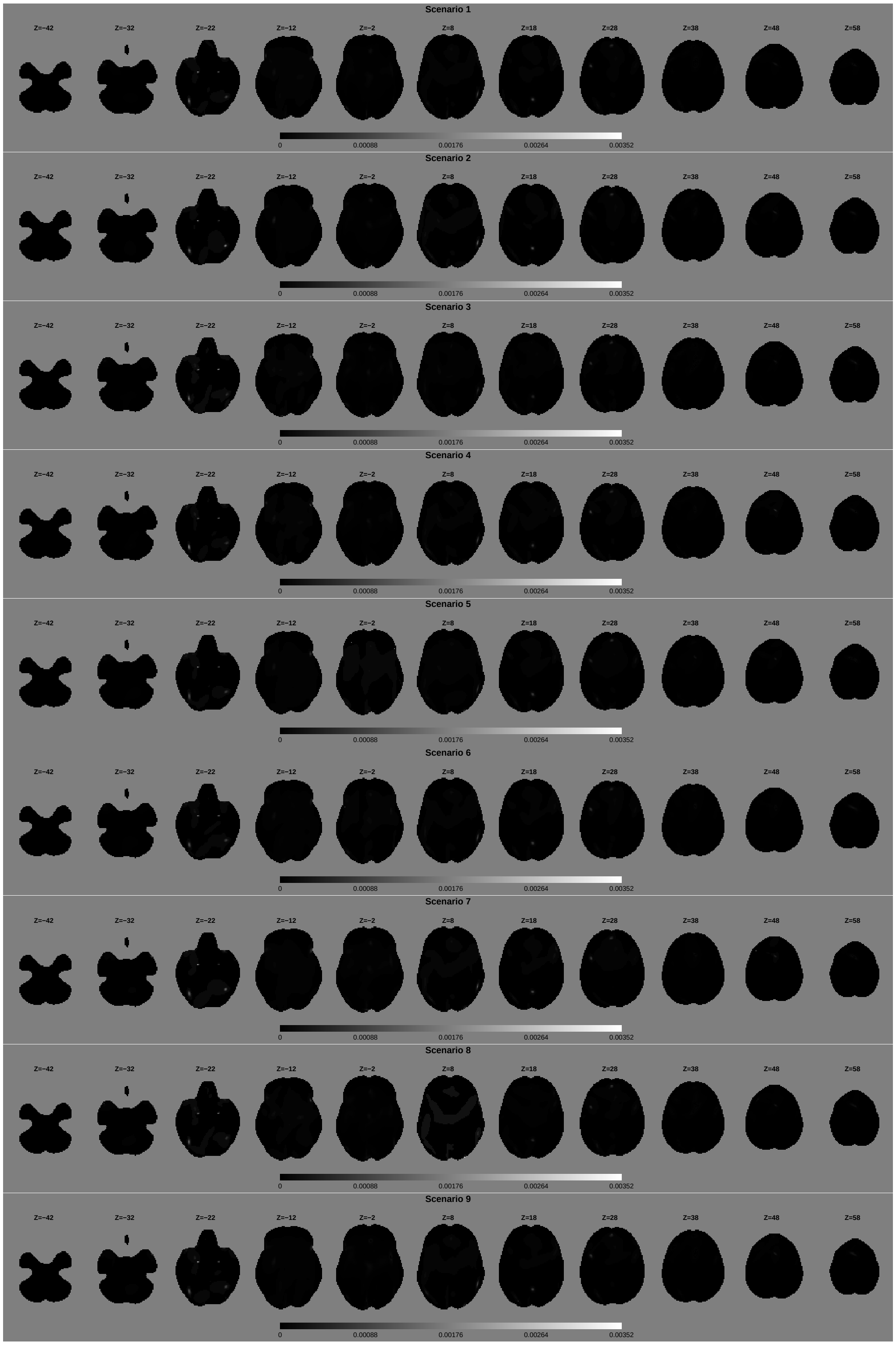


Figure A.2: Sensitivity analysis results: comparisons of the population center intensity on 11 slices of the brain from $Z = -42$mm to $Z = 58$mm.

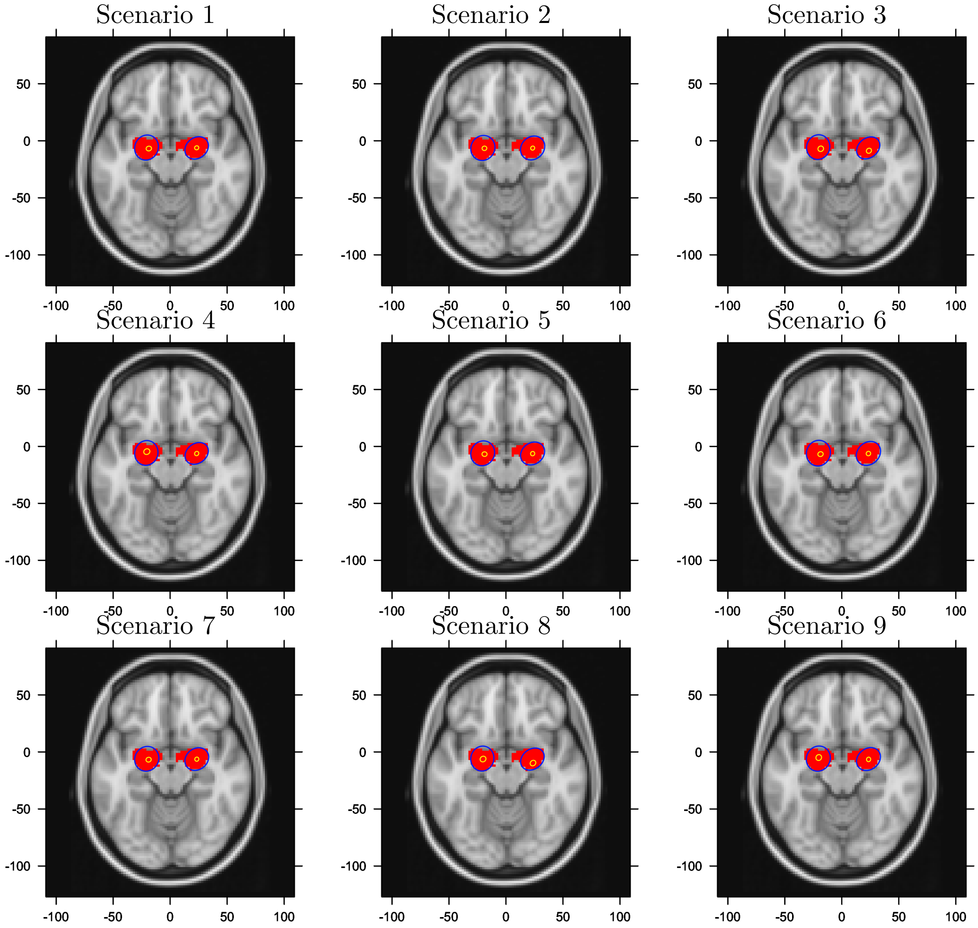


Figure A.3: Sensitivity analysis results: comparisons of the 95% marginal credible ellipses of both population centers (blue circles) and activation centers (yellow circles).

Table A.3: The estimated location (x, y, z), the Euclidean distance in locations between scenario 1 and other scenarios (Dist.), the volume of 95% credible ellipsoid and the volume 95% credible interval of population center as well as activation centers conditional on the population centers located in the amygdalae.

Scenario	Amygdala Region	Location (x, y, z)	Dist.	Volume	95% CI
1	L. Pop. Ctr.	(-18.8, -6.9,-19.8)	0.00	86.80	$[80.1, 94.1]$
	L. Act. Ctr.	(-20.8, -6.0,-19.0)	0.00	12885.80	$[6093.0, 47684.1]$
	R. Pop. Ctr.	(23.2, -6.2,-20.3)	0.00	35.70	$[33.1, 38.6]$
	R. Act. Ctr.	(23.2, -6.3,-19.9)	0.00	8299.00	$[4623.9, 13586.7]$
2	L. Pop. Ctr.	(-19.0, -6.6,-19.8)	0.30	60.00	$[55.4, 64.8]$
	L. Act. Ctr.	(-20.3, -6.4,-19.1)	0.60	12582.80	$[6736.1, 41708.8]$
	R. Pop. Ctr.	(23.4, -6.5,-20.0)	0.40	35.10	$[32.5, 37.9]$
	R. Act. Ctr.	(23.3, -6.4,-19.8)	0.10	9118.20	$[5332.8, 13965.5]$
3	L. Pop. Ctr.	(-18.5, -7.0,-19.8)	0.30	123.90	$[110.7, 142.5]$
	L. Act. Ctr.	(-20.8, -5.6,-19.0)	0.40	11782.70	$[5488.1, 35089.0]$
	R. Pop. Ctr.	(23.7, -6.8,-20.3)	0.80	71.60	$[65.0, 78.6]$
	R. Act. Ctr.	(23.1, -6.3,-19.7)	0.20	8217.70	$[4272.6, 14578.1]$
4	L. Pop. Ctr.	(-18.5, -6.6,-20.3)	0.60	160.80	$[91.4, 193.0]$
	L. Act. Ctr.	(-21.0, -5.8,-19.1)	0.30	12420.30	$[5881.2, 44945.2]$
	R. Pop. Ctr.	(23.2, -6.3,-20.2)	0.20	41.50	$[37.2, 63.4]$
	R. Act. Ctr.	(23.2, -6.2,-19.9)	0.10	8229.20	$[4449.7, 14538.9]$
5	L. Pop. Ctr.	(-18.9, -6.8,-19.8)	0.10	75.40	$[69.4, 82.0]$
	L. Act. Ctr.	(-20.5, -6.1,-19.2)	0.40	12651.90	$[6244.3, 43772.6]$
	R. Pop. Ctr.	(23.3, -6.4,-20.0)	0.40	42.80	$[37.7, 103.2]$
	R. Act. Ctr.	(23.3, -6.3,-19.7)	0.20	8929.50	$[4907.4, 14888.7]$
6	L. Pop. Ctr.	(-18.7, -6.8,-19.7)	0.20	88.20	$[81.1, 95.9]$
	L. Act. Ctr.	(-20.6, -6.0,-19.1)	0.20	13065.90	$[6560.5, 41233.8]$
	R. Pop. Ctr.	(23.3, -6.3,-20.2)	0.20	44.40	$[41.1, 48.0]$
	R. Act. Ctr.	(23.3, -6.2,-19.8)	0.20	9638.40	$[5371.7, 16713.5]$
7	L. Pop. Ctr.	(-19.0, -6.8,-19.8)	0.20	83.10	$[76.4, 90.2]$
	L. Act. Ctr.	(-20.6, -6.0,-19.1)	0.20	12590.50	$[6323.7, 41233.9]$
	R. Pop. Ctr.	(23.3, -6.4,-20.1)	0.30	33.80	$[31.4, 36.5]$
	R. Act. Ctr.	(23.3, -6.3,-19.9)	0.10	8043.00	$[4966.5, 12771.3]$
8	L. Pop. Ctr.	(-18.2, -6.3,-19.5)	0.90	147.90	$[119.0, 252.6]$
	L. Act. Ctr.	(-20.8, -5.8,-19.0)	0.10	12408.30	$[5566.6, 42291.3]$
	R. Pop. Ctr.	(23.6, -6.8,-19.3)	1.20	86.60	$[78.1, 95.0]$
	R. Act. Ctr.	(23.1, -6.3,-19.7)	0.20	8537.70	$[4473.1, 13830.1]$
9	L. Pop. Ctr.	(-18.3, -7.0,-20.3)	0.60	133.50	$[98.8, 151.5]$
	L. Act. Ctr.	(-20.7, -5.9,-19.2)	0.20	13087.00	$[6236.0, 44717.0]$
	R. Pop. Ctr.	(23.4, -6.6,-19.8)	0.70	42.60	$[36.6, 65.4]$
	R. Act. Ctr.	(23.3, -6.3,-19.8)	0.20	8358.00	$[4868.9, 13905.9]$

A.3.2 Sensitivity to Model Specification

Simulation A: We simulate data, or foci, $\mathbf{x}$, the centers of activation regions, $\mathbf{y}$, and the population centers from our model with the following parameters. The mean number of population centers is $\beta|\mathcal{B}| = 5$ and the number of studies is $C = 15$. For each study, we set the mean number of study centers associated with the population centers as $(1 - \sum_{c=1}^{C} \rho_{1c})\theta/C = 1/15$, where $\rho_{1c} = 8/125$ and $\theta = 50$. The mean number of foci, per population center, that do not cluster with a study center is $\sum_{c=1}^{C} \rho_{1c}\theta/C = 4/3$. The mean number of the multiple foci per activation region is $\eta_c = 5$, for all $c = 1, \ldots, C$. The mean number of activation centers that do not cluster is $\epsilon|\mathcal{B}| = 30$. The covariance matrix that describes the variability of activation centers about population centers is set to $\mathbf{\Sigma}_z = 9\boldsymbol{I}$ for all $z \in \mathbf{z}$. The covariance matrix for multiply reported foci that cluster about study centers is $\mathbf{\Psi}_y = \boldsymbol{I}$ for all $y \in \mathbf{y}$. Based on the above settings, the mean number of observed foci is 315. The simulated data are shown in Figure A.4.

The hyper prior values are $E[\beta|\mathcal{B}|] = 2$, $\theta = 100$, $\epsilon|\mathcal{B}| = 50$, $E(\eta_c) = 10$, $E(\rho_{1c}) = 0.03$ for all $c = 1, \cdots, C$, $E[\mathbf{\Sigma}_z] = 4\boldsymbol{I}$, for all $z \in \mathbf{z}$ and $E(\mathbf{\Psi}_y) = \boldsymbol{I}$ for all $y \in \mathbf{y}$.

Results: The posterior mean number of population centers is 5. The estimated posterior marginal intensity function of the activation centers is shown in Figure A.4 from which we can identify the 5 clusters. Also, we can see that the data and the intensity are well matched. The estimated posterior marginal intensity function of the population centers is also shown in Figure A.4. Clearly, the intensity is highly concentrated around the 5 true population locations.

We conclude that if the data are generated from our model, then our method provides very accurate results even when then priors are biased from the truth. Next, we investigate how the proposed method is robust to model mis-specification.

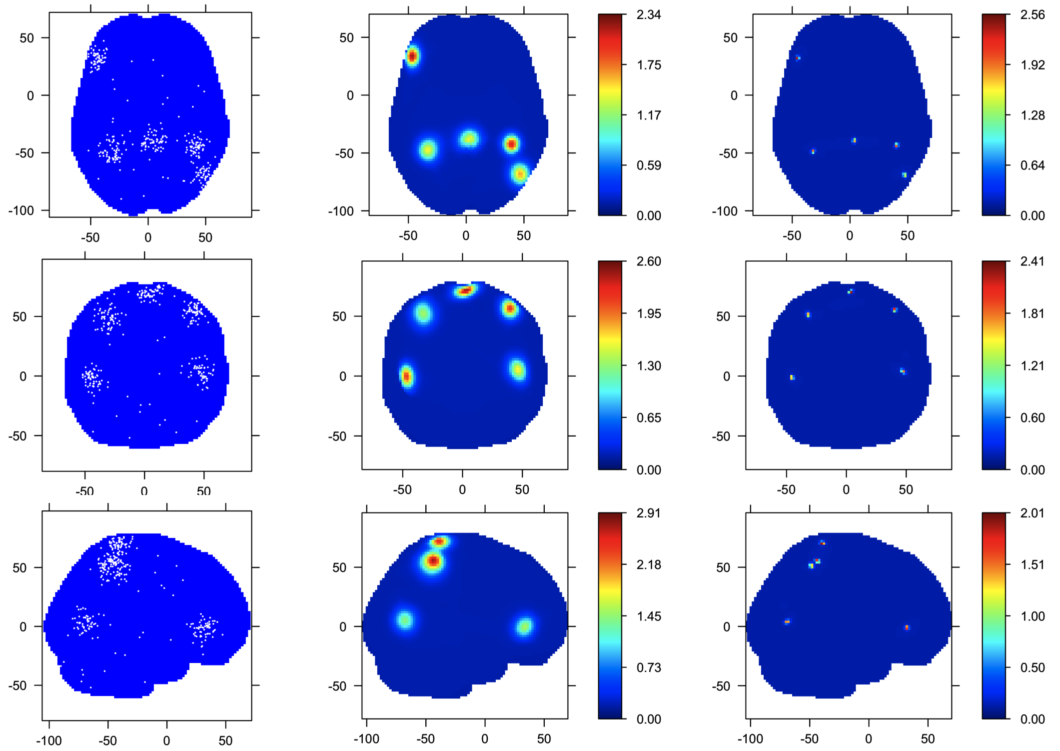


Figure A.4: Simulation A results: the images in the first column show the simulated foci (white points) and the true locations of population centers (red X), projected unto the XY, XZ and YZ plane for the top, middle and bottom, respectively. The images in the middle column show the posterior intensity of the activation centers, integrated over the Z, Y and X directions for the top, middle and bottom, respectively. The images in the last column show the posterior intensity of the population centers, each integrated over one dimension as in the middle column.

Simulation B: We set population centers $\mathbf{z} = \{z_1, z_2\}$, where $z_1 = (22, -6, -18)$ and $z_2 = (-20, -6, -18)$, i.e. the centers of the amygdalae. For each population center $z_i \in \mathbf{z}, i = 1, 2$, we draw 50 foci and 5 study centers from a uniform distribution over spheres with centers z_i and radius $R_{\boldsymbol{\Sigma}} = 5$. For each activation center, $y_j \in \mathbf{y}, j = 1, 2, \cdots, 10$, we draw 10 foci from a uniform distribution over spheres with centers y_j, and radius $R_\Lambda = 1$. The uniform distributions are in direct contrast with the Gaussian distributions assumed in our model. We set the number of studies to $C = 5$ and randomly assign each study two activation centers and 40 foci, for a total of 200 foci. Hyper prior values are $E(\beta|\mathcal{B}|) = 2$, $\epsilon|\mathcal{B}| = 60$, $\theta = 100$, $E(\boldsymbol{\Sigma}_z) = 4\boldsymbol{I}$ for all $z \in \mathbf{z}$, $E(\boldsymbol{\Psi}_y) = \boldsymbol{I}$ for all $y \in \mathbf{y}$, $E(\eta_c) = 10$ and $E(\rho_{1c}) = 0.033$ for all $c = 1, 2, \ldots, C$.

Results are shown in Figure A.5. The estimated posterior marginal intensity of activation centers clearly show there are two activation regions and match the truth well. Also, in Figure A.5, the estimated posterior marginal intensity of the population centers is highly concentrated on the two points. The above results imply that the proposed model is robust to this model mis-specification.

Simulation C: For this simulation, we do not set population centers and study centers. Rather, we directly simulate foci, $\mathbf{x} = \{x_1, \cdots, x_{350}\}$. For $i = 1, 2, \cdots, 300$, we simulate them from the following function:

$$x_i = x_0 + r \begin{pmatrix} \sin(\psi_i)\cos(\varphi_i) \\ \sin(\psi_i)\sin(\varphi_i) \\ \cos(\psi_i) \end{pmatrix} \tag{A.46}$$

where $x_0 = (46, 55, 46)$, $\psi_i \sim \mathrm{U}(-0.5\pi, 0.5\pi)$, $\varphi_i \sim \mathrm{U}(0.25\pi, 0.5\pi)$ and $r \sim \mathrm{U}(20, 25)$. For $i = 301, \cdots, 350$, the x_i are drawn uniformly over the brain and are considered as noise. We then randomly assign the 350 foci to 20 studies. The hyper-prior values

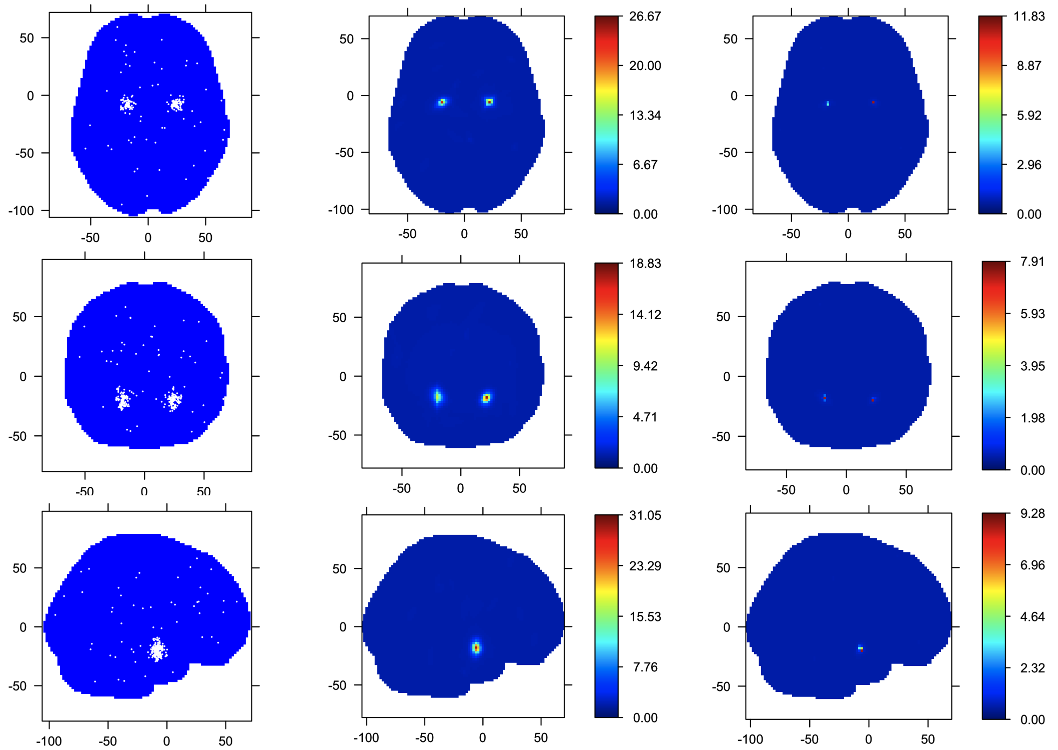


Figure A.5: Simulation B results: The first column shows the simulated foci and the true locations of population centers. The middle column show the posterior intensity of the activation centers, and the last column shows the posterior intensity of the population centers. See Figure 4 for display conventions.

are $E(\beta|\mathcal{B}|) = 5$, $\epsilon|\mathcal{B}| = 60$, $\theta = 50$, $E(\mathbf{\Sigma}_z) = 100\boldsymbol{I}$ for all $z \in \mathbf{z}$, $E(\mathbf{\Psi}_y) = \boldsymbol{I}$ for all $y \in \mathbf{Y}$, $E(\eta_c) = 10$ and $E(\rho_{1c}) = 0.033$ for all $c = 1, 2, \ldots, C$.

Figure A.6 compares the estimated posterior marginal intensity of activation centers with the data. Figure A.6 also compares the posterior intensity of the "population centers"—contrasted against the data as there are no true population centers. Our model picks about 7 population centers. Obviously, in this case, the "population centers" are driven by the data—our model clusters the data about these "population centers" although the data generating mechanism assumes no such centers. The posterior intensity of the activation centers is still well estimated, however, care must be taken in the interpretation of the population level parameters. They exist solely to fit the data.

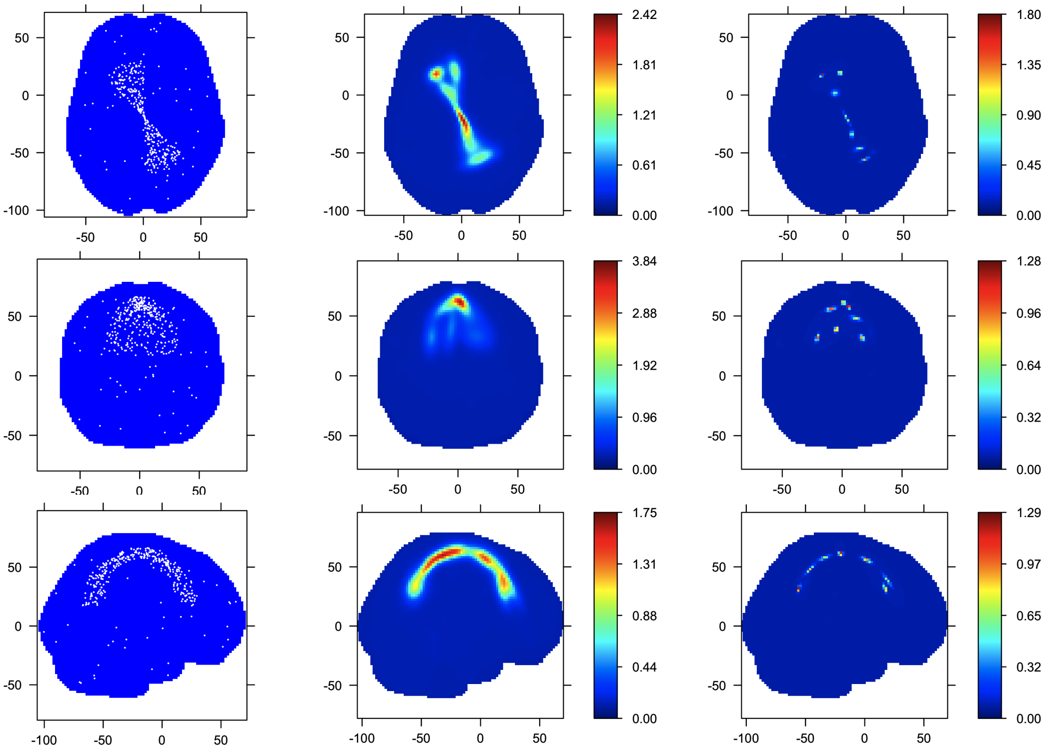


Figure A.6: Simulation C results. The first column shows the simulated foci. The middle column shows the posterior intensity of activation centers, and the last columns shows the posterior intensity of the "population centers". See Figure 4 for display conventions.

A.4 Whole Brain Results

We emphasize here that the model fits intensity functions to the entire brain and not only on ROIs. ROIs, however, are a useful way to summarize the posterior intensity functions. Table A.4 summarizes the posterior mean number of population centers occurring in various ROI as well as the posterior probability that at least one population center occurs in said ROI. Note that most regions are bilateral, specifically the amygdala consists of a pair of symmetric regions. Most ROIs have strong evidence for one or more population centers. ROI's are from the CIC Atlas (Tziortzi et al. 2010), a revised and hierarchical version of the Harvard-Oxford atlas (Desikan et al. 2006).

Table A.4: The posterior expected number of population centers in different ROIs and the posterior probability that at least one population centers is located in the ROI.

Region ($\mathcal{R}$)		$\mathrm{E}[N_{\mathbf{Z}}(\mathcal{R}) \mid \mathbf{X}]$	$\Pr[N_{\mathbf{Z}}(\mathcal{R}) \geq 1 \mid \mathbf{X}]$
Occipital Lobe	Occipital Pole	4.02	>0.999
	Calcarine Cortex	2.00	>0.999
	Cuneus	0.00	0.000
	Lingual Gyrus	1.00	>0.999
	Occipital Fusiform Gyrus	1.00	>0.999
Insular & Temporal Lobe	Insular Cortex	1.00	>0.999
	Anterior Temporal Pole	2.71	>0.999
	Superior Temporal Gyrus	0.41	0.412
	Middle Temporal Gyrus	1.00	>0.999
	Inferior Temporal Gyrus	0.00	0.000
	Parahippocampal Gyrus	0.00	0.000
	Temporal Fusiform Gyrus	1.00	>0.999
	Amygdala	2.05	>0.999
	Hippocampus	0.00	0.000
Frontal Lobe	Precentral Gyrus	0.00	0.000
	DorsoLateral Frontal Cortex	5.00	>0.999
	Medial Frontal Cortex	2.16	>0.999
	Frontal Operculum	0.98	0.982
	Orbitofrontal	4.00	>0.999
	Supplementary Motor Area	0.00	0.000
Cingulate Cortex	Posterior Cingulate Gyrus	0.00	0.000
	Anterior Cingulate	2.00	>0.999
Parietal Lobe	Postcentral Gyrus	0.00	0.000
	Parietal Lobule	0.00	0.000
	Supramarginal Gyrus	0.00	0.000
	Angular Gyrus	0.00	0.000
	Precuneous Cortex	0.00	0.000
	Parietal Operculum Cortex	0.00	0.000
Basal Ganglia	Globus Pallidus	0.00	0.000
	Striatum	1.00	>0.999
Thalamus & Brainstem	Thalamus	0.91	0.913
	Midbrain	0.00	0.000
	Pons	0.00	0.000
	Medulla	0.00	0.000
Cerebellum	Ventrolateral Cerebellum	0.00	0.000
	Medial Cerebellum	0.00	0.000
	Dorsal Cerebellum	2.00	>0.999
Other	Cerebral White Matter	6.00	>0.999
	Lateral Ventricle	0.00	0.000

We also show the posterior standard deviation of activation center intensity function in Figure A.7. Each of 11 slices is a 2mm thick slice, and intensity functions shown in each slice are integrated over the 2mm slice. The axial slices are separated by 10 mm.

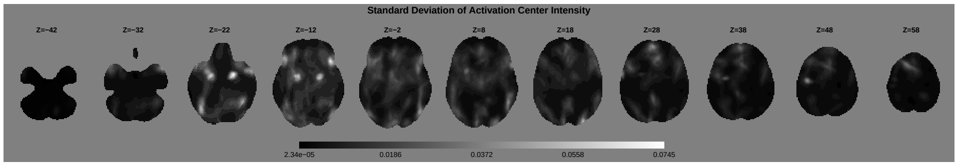


Figure A.7: The posterior standard deviation of the activation center intensity on 11 axial slices of the brain (from $Z = -42$mm to $Z = 58$mm).

APPENDIX B

The Details of Hierarchical Poisson/Gamma Random Fields Model

B.1 Proofs

In this section, we provide the proofs for all the theorems that are presented in the paper. We first introduce the following lemmas.

Lemma 1. *If* $\mathbf{Y} \sim$ *Poisson*$(\mathcal{B}, \Lambda)$*, then for any two regions* $A, B \subset \mathcal{B}$*, we have*

$$\mathrm{Cov}\{N_{\mathbf{Y}}(A), N_{\mathbf{Y}}(B)\} = \mathrm{Var}\{N_{\mathbf{Y}}(A \cap B)\} = \Lambda(A \cap B).$$

Proof. Note that $N_{\mathbf{Y}}(A) = N_{\mathbf{Y}}(A \cap B) + N_{\mathbf{Y}}(A/B)$ and $N_{\mathbf{Y}}(B) = N_{\mathbf{Y}}(A \cap B) + N_{\mathbf{Y}}(B \backslash A)$. By the independent property of the Poisson point process, we have that $N_{\mathbf{Y}}(A)$ and $N_{\mathbf{Y}}(B \backslash A)$ are independent, $N_{\mathbf{Y}}(B)$ and $N_{\mathbf{Y}}(B \backslash A)$ are independent, and $N_{\mathbf{Y}}(B \backslash A)$ and $N_{\mathbf{Y}}(A \backslash B)$ are independent. Thus,

$$\begin{aligned} \mathrm{Cov}\{N_{\mathbf{Y}}(A), N_{\mathbf{Y}}(B)\} &= \mathrm{Cov}\{N_{\mathbf{Y}}(A \cap B) + N_{\mathbf{Y}}(A/B), N_{\mathbf{Y}}(A \cap B) + N_{\mathbf{Y}}(B/A)\} \\ &= \mathrm{Cov}\{N_{\mathbf{Y}}(A \cap B), N_{\mathbf{Y}}(A \cap B)\} = \mathrm{Var}\{N_{\mathbf{Y}}(A \cap B)\} = \Lambda(A \cap B). \end{aligned}$$

□

Lemma 2. *Let* $\Gamma(d\mathrm{x}) \sim \mathcal{GRF}\{\alpha(d\mathrm{x}), \beta\}$*, and let* $f_1(\mathrm{x})$ *and* $f_2(\mathrm{x})$ *be measurable functions on* $\mathcal{B}$*, Then*

$$\mathrm{Cov}\left\{\int_{\mathcal{B}} f_1(\mathrm{x}) d\Gamma(d\mathrm{x}), \int_{\mathcal{B}} f_2(\mathrm{x}) d\Gamma(d\mathrm{x})\right\} = \frac{1}{\beta^2} \int_{\mathcal{B}} f_1(\mathrm{x}) f_2(\mathrm{x}) \alpha(d\mathrm{x})$$

Proof. We start with the case that $f_1(\mathrm{x})$ and $f_2(\mathrm{x})$ are simple functions on $\mathcal{B}$, i.e. for $m = 1, \ldots, M$ and $n = 1, \ldots, M$, there exist numbers $a_m, b_n \in \mathcal{R}$ and disjointed sets A_m and B_n with $\mathcal{B} = \bigcup_{m=1}^{M} A_m = \bigcup_{n=1}^{N} B_n$ such that $f_1(\mathrm{x}) = \sum_{m=1}^{M} a_m \delta_{A_m}(\mathrm{x})$ and $f_2(\mathrm{x}) = \sum_{n=1}^{N} b_n \delta_{B_n}(\mathrm{x})$. Then $\int_{\mathcal{B}} f_1(\mathrm{x})\Gamma(d\mathrm{x}) = \sum_{m=1}^{M} a_m \Gamma(A_m)$ and $\int_{\mathcal{B}} f_2(\mathrm{x})\Gamma(d\mathrm{x}) = \sum_{n=1}^{N} b_n \Gamma(B_n)$. Thus,

$$
\begin{aligned}
&\mathrm{E}\left\{\int_{\mathcal{B}} f_1(\mathrm{x}) d\Gamma(d\mathrm{x}) \times \int_{\mathcal{B}} f_2(\mathrm{x}) d\Gamma(d\mathrm{x})\right\} = E\left\{\sum_{m=1}^{M} a_m \Gamma(A_m) \times \sum_{n=1}^{N} b_n \Gamma(B_n)\right\} \\
&= E\left\{\sum_{m=1}^{M}\sum_{n=1}^{N} a_m \Gamma(A_m \cap B_n) \times \sum_{n=1}^{N}\sum_{m=1}^{M} b_n \Gamma(A_m \cap B_n)\right\} \\
&= \sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{m'=1}^{M}\sum_{n'=1}^{N} a_m b_{n'} E\left\{\Gamma(A_m \cap B_n)\Gamma(A_{m'} \cap B_{n'})\right\} \\
&= \sum_{m=1}^{M}\sum_{n=1}^{N} a_m b_n E\{\Gamma^2(A_m \cap B_n)\} \\
&\qquad + \sum_{(m,n)\neq(m',n')} a_m b_{n'} E\{\Gamma(A_m \cap B_n)\} E\{\Gamma(A_{m'} \cap B_{n'})\} \\
&= \sum_{m=1}^{M}\sum_{n=1}^{N} a_m b_n \left\{\frac{\alpha(A_m \cap B_n)}{\beta^2} + \frac{\alpha^2(A_m \cap B_n)}{\beta^2}\right\} \\
&\qquad + \sum_{(m,n)\neq(m',n')} a_m b_{n'} \frac{\alpha(A_m \cap B_n)}{\beta} \frac{\alpha(A_{m'} \cap B_{n'})}{\beta} \\
&= \frac{1}{\beta^2}\sum_{m=1}^{M}\sum_{n=1}^{N} a_m b_n \alpha(A_m \cap B_n) \\
&\qquad + \frac{1}{\beta^2}\sum_{m=1}^{M}\sum_{n=1}^{N}\sum_{m'=1}^{M}\sum_{n'=1}^{N} a_m b_{n'} \alpha(A_m \cap B_n)\alpha(A_{m'} \cap B_{n'}) \\
&= \frac{1}{\beta^2}\int_{\mathcal{B}} f_1(\mathrm{x}) f_2(\mathrm{x})\alpha(d\mathrm{x}) + \frac{1}{\beta^2}\int_{\mathcal{B}} f_1(\mathrm{x})\alpha(d\mathrm{x}) \int_{\mathcal{B}} f_2(\mathrm{x})\alpha(d\mathrm{x}).
\end{aligned}
$$

Furthermore,

$$
\begin{aligned}
&\mathrm{Cov}\left\{\int_{\mathcal{B}} f_1(\mathrm{x}) d\Gamma(d\mathrm{x}), \int_{\mathcal{B}} f_2(\mathrm{x}) d\Gamma(d\mathrm{x})\right\} \\
&= E\left\{\int_{\mathcal{B}} f_1(\mathrm{x}) d\Gamma(d\mathrm{x}) \int_{\mathcal{B}} f_2(\mathrm{x}) d\Gamma(d\mathrm{x})\right\} - \frac{1}{\beta^2}\int_{\mathcal{B}} f_1(\mathrm{x})\alpha(d\mathrm{x}) \int_{\mathcal{B}} f_2(\mathrm{x})\alpha(d\mathrm{x}) \\
&= \frac{1}{\beta^2}\int_{\mathcal{B}} f_1(\mathrm{x}) f_2(\mathrm{x})\alpha(d\mathrm{x}).
\end{aligned}
$$

For general measurable functions, $f_1(\mathrm{x})$ and $f_2(\mathrm{x})$, a routine passage to the limit, refining the partition A_m and B_n, completes the proof. $\square$

B.1.1 Proof of Theorem 1

First, we note that given G_0, τ and σ_j^2, the conditional expectation of $N_{\mathbf{Y}_j}(A)$ is

$$\begin{aligned}
&E\{N_{\mathbf{Y}_j}(A) \mid G_0, \tau, \sigma_j^2\} = E_{G_j}\{E\{N_{\mathbf{Y}_j}(A) \mid G_j\} \mid G_0, \tau, \sigma_j^2\} \\
&\quad = E_{G_j}\{\Lambda_j(A) \mid G_0, \tau, \sigma_j^2\} = E_{G_j}\left\{\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) G_j(d\mathrm{x}) \mid G_0, \tau\right\} \\
&\quad = \int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) E\left\{G_j(d\mathrm{x}) \mid G_0, \tau\right\} = \frac{1}{\tau}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) G_0(d\mathrm{x}).
\end{aligned}$$

Also, the conditional covariance between $N_{\mathbf{Y}_j}(A)$ and $N_{\mathbf{Y}_j}(B)$ is

$$\begin{aligned}
&\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_j}(B) \mid G_0, \tau, \sigma_j^2\} \\
&\quad = E_{G_j}\{\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_j}(B) \mid G_j\} \mid G_0, \tau, \sigma_j^2\} \\
&\qquad\qquad + \mathrm{Cov}_{G_j}\{E\{N_{\mathbf{Y}_j}(A) \mid G_j\}, E\{N_{\mathbf{Y}_j}(B) \mid G_j\} \mid G_0, \tau, \sigma_j^2\} \\
&\quad = E\{\Lambda_j(A \cap B) \mid G_0, \tau\} + \mathrm{Cov}\{\Lambda_j(A), \Lambda_j(B) \mid G_0, \tau, \sigma_j^2\} \\
&\quad = \frac{1}{\tau}\int_{\mathcal{B}} K_{\sigma_j^2}(A \cap B, \mathrm{x}) G_0(d\mathrm{x}) + \frac{1}{\tau^2}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) K_{\sigma_j^2}(B, \mathrm{x}) G_0(d\mathrm{x}).
\end{aligned}$$

Now, given σ_j^2, α, β, within type j, for any $A \subseteq \mathcal{B}$,

$$\begin{aligned}
&E\{N_{\mathbf{Y}_j}(A) \mid \sigma_j^2, \tau, \alpha, \beta\} = E_{G_0}\{E\{N_{\mathbf{Y}_j}(A) \mid G_0, \sigma_j^2, \tau\} \mid \sigma_j^2, \tau, \alpha, \beta\} \\
&\quad = E_{G_0}\left\{\frac{1}{\tau}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) G_0(d\mathrm{x}) \mid \sigma_j^2, \tau, \alpha, \beta\right\} \\
&\quad = \frac{1}{\tau}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) E\{G_0(d\mathrm{x}) \mid \alpha, \beta\} = \frac{1}{\tau\beta}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x}) \alpha(d\mathrm{x}).
\end{aligned}$$

Within type j, the conditional covariance between $N_{\mathbf{Y}_j}(A)$ and $N_{\mathbf{Y}_j}(B)$ is

$$
\begin{aligned}
&\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_j}(B) \mid \sigma_j^2, \tau, \alpha, \beta\} \\
&\quad= E_{G_0}\{\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_j}(B) \mid G_0, \tau, \sigma_j^2\} \mid \sigma_j^2, \tau, \alpha, \beta\} \\
&\qquad\qquad +\mathrm{Cov}_{G_0}\{E\{N_{\mathbf{Y}_j}(A) \mid G_0, \tau, \sigma_j^2\}, E\{N_{\mathbf{Y}_j}(B) \mid G_0, \tau, \sigma_j^2\} \mid \tau, \sigma_j^2, \alpha, \beta\} \\
&\quad= E_{G_0}\left\{\frac{1}{\tau}\int_{\mathcal{B}} K_{\sigma_j^2}(A\cap B, \mathrm{x})G_0(d\mathrm{x}) + \frac{1}{\tau^2}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x})K_{\sigma_j^2}(B, \mathrm{x})G_0(d\mathrm{x})\right\} \\
&\qquad\qquad +\mathrm{Cov}\left\{\frac{1}{\tau}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x})G_0(d\mathrm{x}), \frac{1}{\tau}\int_{\mathcal{B}} K_{\sigma_j^2}(B, \mathrm{x})G_0(d\mathrm{x}) \mid \tau, \sigma_j^2, \alpha, \beta\right\} \\
&\quad= \frac{1}{\tau\beta}\int_{\mathcal{B}} K_{\sigma_j^2}(A\cap B, \mathrm{x})\alpha(d\mathrm{x}) + \frac{1}{\tau^2\beta}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x})K_{\sigma_j^2}(B, \mathrm{x})\alpha(d\mathrm{x}) \\
&\qquad\qquad +\frac{1}{\tau^2\beta^2}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x})K_{\sigma_j^2}(B, \mathrm{x})\alpha(d\mathrm{x}) \\
&\quad= \frac{1}{\tau\beta}\int_{\mathcal{B}} K_{\sigma_j^2}(A\cap B, \mathrm{x})\alpha(d\mathrm{x}) + \frac{1+\beta}{\tau^2\beta^2}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x})K_{\sigma_j^2}(B, \mathrm{x})\alpha(d\mathrm{x}).
\end{aligned}
$$

For $j \neq k$, the conditional covariance between $N_{\mathbf{Y}_j}(A)$ and $N_{\mathbf{Y}_k}(B)$ is

$$
\begin{aligned}
&\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_k}(B) \mid \sigma_j^2, \sigma_k^2, \tau, \alpha, \beta\} \\
&\quad= E_{G_0}\{\mathrm{Cov}\{N_{\mathbf{Y}_j}(A), N_{\mathbf{Y}_k}(B) \mid G_0, \sigma_j^2, \sigma_k^2, \tau\} \mid \sigma_j^2, \sigma_k^2, \tau, \alpha, \beta\} \\
&\qquad\quad +\mathrm{Cov}_{G_0}\{E\{N_{\mathbf{Y}_j}(A) \mid G_0, \sigma_j^2, \tau\}, E\{N_{\mathbf{Y}_k}(B) \mid G_0, \sigma_k^2, \tau\} \mid \sigma_k^2, \sigma_j^2, \tau, \alpha, \beta\} \\
&\quad= \frac{1}{\tau^2\beta^2}\int_{\mathcal{B}} K_{\sigma_j^2}(A, \mathrm{x})K_{\sigma_k^2}(B, \mathrm{x})\alpha(d\mathrm{x}).
\end{aligned}
$$

B.1.2 Proof of Theorem 3

Proof. For any $M > 0$, the conditional expected truncated error given σ_j^2, τ, β, and $\{\nu_m, \theta_m\}_{m=1}^M$ is

$$
\begin{aligned}
&E\left[\Lambda_j(A) - \Lambda_j^M(A) \mid \sigma_j^2, \tau, \{\nu_m, \theta_m\}_{m=1}^M\right] \\
&\quad = \frac{1}{\tau}\sum_{m=M+1}^{\infty} E\{\nu_m \mid \{\nu_m\}_{m=1}^M\} K_{\sigma_j^2}(A, \theta_m) \\
&\quad \leq \frac{1}{\tau}\sum_{m=M+1}^{\infty} E\{\nu_m \mid \{\nu_m\}_{m=1}^M, \beta\} = \frac{1}{\tau}\sum_{m=1}^{\infty} E\{\nu_{m+M} \mid \{\nu_m\}_{m=1}^M, \beta\} \\
&\quad = \frac{1}{\tau\beta}\sum_{m=1}^{\infty} E\left\{E_1^{-1}(\zeta_{m+M}/\alpha(\mathcal{B})) \mid \{\nu_m\}_{m=1}^M\right\} \quad (\text{let } \zeta'_m = \zeta_{m+M} - \zeta_M) \\
&\quad = \frac{1}{\tau\beta}\sum_{m=1}^{\infty} E\left\{E_1^{-1}((\zeta'_m + \zeta_M)/\alpha(\mathcal{B})) \mid \{\nu_m\}_{m=1}^M\right\} (\text{note that } \zeta'_m \sim \text{Gamma}(m, 1)) \\
&\quad = \frac{1}{\tau\beta}\sum_{m=1}^{\infty} \int_0^{\infty} E_1^{-1}((s + \zeta_M)/\alpha(\mathcal{B}))\frac{s^{m-1}}{\Gamma(m)}\exp(-s)ds \\
&\quad = \frac{1}{\tau\beta}\int_0^{\infty} E_1^{-1}((s + \zeta_M)/\alpha(\mathcal{B}))\sum_{m=1}^{\infty}\frac{s^{m-1}}{\Gamma(m)}\exp(-s)ds \\
&\quad = \frac{1}{\tau\beta}\int_0^{\infty} E_1^{-1}((s + \zeta_M)/\alpha(\mathcal{B}))ds \\
&\quad = \frac{1}{\tau\beta}\left[1 - \exp\{-E_1^{-1}(\zeta_M/\alpha(\mathcal{B}))\}\right] \\
&\quad \leq \frac{E_1^{-1}(\zeta_M/\alpha(\mathcal{B}))}{\tau\beta} \leq \frac{1}{(\exp(\zeta_M/\alpha(\mathcal{B})) - 1)\tau\beta}.
\end{aligned}
$$

Taking the expectation with respect to ζ_M on both sides of the above inequality,

$$
\begin{aligned}
&E\left[\Lambda_j(A) - \Lambda_j^M(A) \mid \tau, \beta\right] \\
&\quad \leq \frac{1}{\tau\beta\Gamma(M)}\int_0^{\infty}\frac{s^{M-1}\exp(-s)}{(\exp(s/\alpha(\mathcal{B})) - 1)}ds \\
&\quad \leq \frac{\alpha(\mathcal{B})}{\tau\beta\Gamma(M)}\int_0^{\infty} s^{M-2}\exp(-s)ds = \frac{\alpha(\mathcal{B})}{\tau\beta(M-1)}.
\end{aligned}
$$

when $\alpha(\mathcal{B}) = 1$, we have a more accurate upper bound

$$E\left[\Lambda_j(A) - \Lambda_j^M(A) \mid \tau, \beta\right] \leq \frac{\zeta(M) - 1}{\tau\beta}.$$

where $\zeta(M) = \sum_{k=1}^{\infty} k^{-M}$. Therefore, for any $\epsilon > 0$, take $M_\epsilon = \lfloor a(\mathcal{B})/\tau\beta\epsilon \rfloor + 1$, where $\lfloor x \rfloor = n$ if $n \leq x < n+1$ and n is an integer number. This completes the proof. $\square$

B.1.3 Proof of Theorem 4

Proof. Write $a_{j,l} = \sum_{m=1}^{\tilde{N}} m I_{\theta_m}(\mathrm{x}_{j,l})$, then we have $I_m(a_{j,l}) = I_{\theta_m}(\mathrm{x}_{j,l})$. Note that the θ_m are distinct, $\sum_{m=1}^{M} I_{\theta_m}(\mathrm{x}_{j,l}) = 1 = \sum_{m=1}^{M} I_m(a_{j,l})$. Thus,

$$\sum_{m=1}^{M} \mu_{j,m} I_{\theta_m}(\mathrm{x}_{j,l}) = \sum_{m=1}^{M} \mu_{j,m} I_m(a_{j,l}) = \mu_{j a_{j,l}}. \tag{B.1}$$

Furthermore,

$$\prod_{l=1}^{m_{i,j}} \mu_{j,a_{j,l}} = \prod_{l=1}^{m_{i,j}} \mu_{j,a_{j,l}}^{\sum_{m=1}^{M} I_m(a_{j,l})} = \prod_{l=1}^{m_{i,j}} \prod_{m=1}^{M} \mu_{j,a_{j,l}}^{I_m(a_{j,l})} = \prod_{m=1}^{M} \mu_{j,m}^{b_{j,m}}, \tag{B.2}$$

where $b_{j,m} = \sum_{l=1}^{m_{i,j}} I_{\theta_m}(\mathrm{x}_{j,l})$. From (3.14), (B.1) and (B.2), the conditional distribution of $\mu_{j,m}$ is

$$\pi(\mu_{j,m} \mid \cdot) \propto \mu_{j,m}^{\nu_m + b_{j,m} - 1} \exp\{-(K_\sigma(\mathcal{B}, \mathrm{x}) + \theta_m)\mu_{j,m}\}$$

which completes the proof. $\square$

B.1.4 Proof of Theorem 5

Proof. It is straightforward to obtain (3.19) from (3.14). Now we show $\pi(\nu_m \mid \cdot)$ for $m = 1, \ldots, M$, are log concave functions. Define $g(\nu_m) = \nu_m \log(c_m) - \log(\nu_m) - \log\{\Gamma(\nu_m)\}$. By the Taylor expansion of $\log(\Gamma(x))$ (Boros and Moll 2004),

$$\log(\Gamma(x)) = -\gamma x - \log(x) - \sum_{k=1}^{\infty} \left[\frac{x}{k} - \log\left(1 + \frac{x}{k}\right)\right],$$

where γ is a constant. Thus,

$$g(\nu_m) = \nu_m[\gamma + \log(c_m)] - \sum_{k=1}^{\infty} \left[\frac{\nu_m}{k} - \log\left(1 + \frac{\nu_m}{k}\right)\right].$$

Furthermore, the second order derivative of $\log[\pi(\nu_m \mid \cdot)]$, for $m = 1, \ldots, M-1$, is

$$[\log(\pi(\nu_m \mid \cdot))]'' = g''(\nu_m) = -\sum_{k=1}^{\infty}\left[\frac{1}{(\nu_m^2 + k)^2}\right] < 0,$$

and

$$\begin{aligned}[\log\{\pi(\nu_M \mid \cdot)\}]'' &= \beta^2 E_1''(\beta\nu_M) + g''(\nu_M) \\ = \quad & -\frac{1}{\nu_M^2}(\beta\nu_M + 1)\exp(-\beta\nu_M) - \sum_{k=1}^{\infty}\left[\frac{1}{(\nu_M^2 + k)^2}\right] < 0.\end{aligned}$$

This implies that $\pi(\nu_m \mid \cdot)$, for $m = 1, \ldots, M$, are log-concave. □

B.2 PGRF for multiple spatial point patterns

The PGRF model (Wolpert and Ickstadt 1998a) fits the intensity function of a single spatial point pattern. We extend the PGRF model on two fronts: we allow multi-type point patterns, and within each type, we allow multiple, independent realizations. Specifically, we assume that $\mathbf{y}_{i,j}$, for $i = 1, \ldots, n_j$, are independent realizations of a Poisson point process $\mathbf{Y}_j$ with intensity measure $\Lambda_j(d\mathrm{y})$ which, in turn, is the convolution of a random measure $G_j(d\mathrm{x})$ with a kernel measure $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x})$:

$$\begin{aligned}\mathbf{Y}_j \mid \Lambda_j &\overset{i.i.d.}{\sim} \text{Poisson}\{\mathcal{B}, \Lambda_j(d\mathrm{y})\}, \quad \Lambda_j(d\mathrm{y}) = \int_{\mathcal{B}} K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}) G_j(d\mathrm{x}) \\ G_j(d\mathrm{x}) &\sim \mathcal{GRF}\{G_0(d\mathrm{x}), \tau\}. \end{aligned} \tag{B.3}$$

In this section, we assume that the different groups of spatial point patterns are independent of one another, then the base measure $G_0(d\mathrm{x})$ in model (B.3) is considered as a prior parameter and assumed to be fixed and known. This implies that we can separately fit the model within each group to make inference on the random intensity measure $\Lambda_j(d\mathrm{x})$ and $G_j(d\mathrm{x})$ for $j = 1, \ldots, J$ in model (B.3). When $G_0(d\mathrm{x})$ is fixed, we refer the model (B.3) as the IPGRF, which extends the original

PGRF model (Wolpert and Ickstadt 1998a) to fit the multiple independent realizations of a Poisson point process. In particular, we consider independent realizations, i.e. $\{(\mathbf{x}_{\cdot,j}, \mathbf{y}_{\cdot,j})\} = \{\mathbf{x}_{i,j}, \mathbf{y}_{i,j}\}_{i=1}^{m_j}$ of $(\mathbf{X}_j, \mathbf{Y}_j)$, for $j = 1, \ldots, J$, in model (3.8). The following theorem suggests a posterior simulation algorithm for $G_j(d\mathrm{x})$.

Theorem 6. *The conditional distribution of* $G_j(d\mathrm{x})$*, given* τ*,* σ_j^2*,* G_0 *and* $(\mathbf{y}_{\cdot j}, \mathbf{x}_{\cdot j})$*, for* $j = 1, \ldots, J$*, is given by*

$$\begin{aligned} & [G_j(d\mathrm{x}) \mid \tau, \sigma_j^2, G_0, \mathbf{y}_{\cdot j}, \mathbf{x}_{\cdot j}] \\ & \quad \sim \mathcal{GRF}\left\{ G_0(d\mathrm{x}) + \sum_{i=1}^{n_j} N_{\mathbf{x}_{i,j}}(d\mathrm{x}), \tau + n_j K_{\sigma_j^2}(\mathcal{B}, \mathrm{x}) \right\}. \end{aligned} \tag{B.4}$$

Proof. First let $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x})$, for $j = 1, \ldots, J$, be simple functions on $\mathcal{B}$ with respect to x, i.e. there exists a set of partitions $\{A_{jm}\}_{m=1}^{M_j}$, for $j = 1, \ldots, J$, of $\mathcal{B}$, such that $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}_1) \equiv K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x}_2)$ for any $\mathrm{x}_1, \mathrm{x}_2 \in A_{jm}$.

Let $\chi_{jm} = K_{\sigma_j^2}(\mathcal{B}, \mathrm{x})$ for $\mathrm{x} \in A_{jm}$, $\kappa_{jm}^{il} = k_{\sigma_j^2}(\mathrm{y}_{j,l}, \mathrm{x})$ for $\mathrm{x} \in A_{jm}$, $\gamma_{jm} = G_j(A_m)$ and $g_m = G_0(A_m)$. This implies that $\Lambda_j(\mathcal{B}) = \sum_{m=1}^{M_j} \chi_{jm}\gamma_m$. In model (3.8), the joint distribution of $(\mathbf{x}_{\cdot,j}, \mathbf{y}_{\cdot,j}), \{G_j\}_{j=1}^J$ given G_0, $\boldsymbol{\sigma}$ and τ is

$$\begin{aligned} & \pi(\{\mathbf{x}_{\cdot,j}, \mathbf{y}_{\cdot,j}\}_{j=1}^J, \{G_j\}_{j=1}^J \mid G_0, \{\sigma_j^2\}_{j=1}^J, \tau) \\ & \quad = \pi(\{(\mathbf{x}_{\cdot,j}, \mathbf{y}_{\cdot,j})\}_{j=1}^J, \mid \{G_j\}_{j=1}^J, \{\sigma_j^2\}_{j=1}^J) \pi(\{G_j\}_{j=1}^J \mid G_0, \tau) \\ & \quad = \prod_{j=1}^J \left\{ \prod_{i=1}^{n_j} \left[\exp\left(|\mathcal{B}| - \sum_{m=1}^{M_j} \chi_{jm}\gamma_{jm} \right) \prod_{l=1}^{m_{i,j}} \sum_{m=1}^{M_j} k_{jm}^{il} \gamma_{jm} \delta_{\mathrm{x}_{j,l}}(A_{jm}) \right] \right\} \\ & \qquad\qquad \times \prod_{j=1}^J \left\{ \prod_{m=1}^{M_j} \frac{\gamma_{jm}^{g_m - 1} \tau^{g_m}}{\Gamma(g_m)} \exp\left(-\tau\gamma_{jm}\right) \right\}. \end{aligned}$$

This implies that the posterior distribution of G_j given G_0, $\boldsymbol{\sigma}$ and τ,

$$\begin{aligned} & \pi(G_j \mid G_0, \boldsymbol{\sigma}, \tau, \mathbf{y}, \mathbf{x}) \\ & \quad \propto \exp\left(-\sum_{m=1}^{M_j} (n_j\chi_{jm} + \tau)\gamma_{jm} \right) \prod_{i=1}^{n_j} \prod_{l=1}^{m_{i,j}} \sum_{m=1}^{M_j} k_{jm}^{il} \gamma_{jm} \delta_{\mathrm{x}_{j,l}}(A_{jm}) \cdot \prod_{m=1}^{M_j} \gamma_{jm}^{g_m - 1}. \end{aligned}$$

Note that $\sum_{m=1}^{M_j} \delta_{\mathrm{x}_{j,l}}(A_{jm}) = 1$ for any i, j and l. Let $t_{j,l} = \sum_{m=1}^{M_j} m\delta_{\mathrm{x}_{j,l}}(A_{jm})$, then $I_m(t_{j,l}) = \delta_{\mathrm{x}_{j,l}}(A_{jm})$. This implies that

$$\sum_{m=1}^{M_j} k_{jm}^{il}\gamma_{jm}\delta_{\mathrm{x}_{j,l}}(A_{jm}) = \sum_{m=1}^{M_j} k_{jm}^{il}\gamma_{jm}I_m(t_{j,l}) = k_{jt_{j,l}}^{il}\gamma_{jt_{j,l}},$$

and

$$\begin{aligned}
\prod_{i=1}^{n_j}\prod_{l=1}^{m_{i,j}} k_{jt_{j,l}}^{il}\gamma_{jt_{j,l}} &= \prod_{i=1}^{n_j}\prod_{l=1}^{m_{i,j}} \{k_{jt_{j,l}}^{il}\gamma_{jt_{j,l}}\}^{\sum_{m=1}^{M_j} I_m(t_{j,l})} \\
&= \prod_{i=1}^{n_j}\prod_{l=1}^{m_{i,j}}\prod_{m=1}^{M_j} \{k_{jt_{j,l}}^{il}\gamma_{jt_{j,l}}\}^{I_m(t_{j,l})} = \prod_{m=1}^{M_j}\prod_{i=1}^{n_j}\prod_{l=1}^{m_{i,j}} \{k_{jm}^{il}\gamma_{jm}\}^{I_m(t_{j,l})} \\
&= \prod_{m=1}^{M_j} \{k_{jm}^{il}\gamma_{jm}\}^{\sum_{i=1}^{n_j}\sum_{l=1}^{m_{i,j}} I_m(t_{j,l})} = \prod_{m=1}^{M_j} \{k_{jm}^{il}\gamma_{jm}\}^{\sum_{i=1}^{n_j}\sum_{l=1}^{m_{i,j}} \delta_{\mathrm{x}_{j,l}}(A_{jm})} \\
&= \prod_{m=1}^{M_j} \{k_{jm}^{il}\gamma_{jm}\}^{\sum_{i=1}^{n_j} N_{\mathrm{x}_{ij}}(A_{jm})},
\end{aligned}$$

which in turn implies that

$$\pi(G_j \mid G_0, \boldsymbol{\sigma}, \tau, \mathbf{y}, \mathbf{x}) \propto \prod_{m=1}^{M_j} \left\{ \exp\left[-(n_j\chi_{jm} + \tau)\gamma_{jm}\right] \gamma_{jm}^{\sum_{i=1}^{n_j} N_{\mathrm{x}_{ij}}(A_{jm}) + g_m - 1} \right\}.$$

Let $\phi_j(\mathrm{x})$ be a nonnegative and simple function with constant value ϕ_{jm} on each A_{jm}. Then the moment generating function is $G_j[\phi] = \sum_{m=1}^{M_j} \phi_{jm}\gamma_{jm}$. Also,

$$\begin{aligned}
&E\{\exp(-G_j[\phi]) \mid G_0, \boldsymbol{\sigma}, \tau, \mathbf{y}, \mathbf{x}\} \\
&\quad\propto \prod_{m=1}^{M_j} \left\{ \int_0^\infty \exp\left(-(\phi_{jm} + n_j\chi_{jm} + \tau)\gamma_{jm}\right) \gamma_{jm}^{\sum_{i=1}^{n_j} N_{\mathrm{x}_{ij}}(A_{jm}) + g_m - 1} d\gamma_{jm} \right\} \\
&\quad\propto \prod_{m=1}^{M_j} \{\phi_{jm} + n_j\chi_{jm} + \tau\}^{-\sum_{i=1}^{n_j} N_{\mathrm{x}_{ij}}(A_{jm}) - g_m}
\end{aligned}$$

The normalizing constant can be obtained by $\exp\{-G_j(0)\} = 1$. Then the Laplace exponent of G_j given G_0, $\boldsymbol{\sigma}$, τ, $\mathbf{y}$ and $\mathbf{x}$ is,

$$\begin{aligned}
&-\log\{E\{\exp(-G_j[\phi]) \mid G_0, \boldsymbol{\sigma}, \tau, \mathbf{y}, \mathbf{x}\}\} \\
&\quad = \log \prod_{m=1}^{M_j} \left\{ \frac{\phi_{jm} + n_j\chi_{jm} + \tau}{n_j\chi_{jm} + \tau} \right\}^{\sum_{i=1}^{n_j} N_{\mathrm{x}_{ij}}(A_{jm}) + g_m} \\
&\quad = \sum_{m=1}^{M_j} \log\{1 + \phi_{jm}(n_j\chi_{jm} + \tau)^{-1}\} \times \left\{ \sum_{i=1}^{n_j} N_{\mathrm{x}_{ij}}(A_{jm}) + g_m \right\} \\
&\quad = \int_{\mathcal{B}} \log\{1 + \phi(\mathrm{x})(n_j K_{\sigma_j^2}(\mathcal{B}, \mathrm{x}) + \tau)^{-1}\} \left(\sum_{i=1}^{n_j} N_{\mathrm{x}_{ij}} + G_0 \right)(d\mathrm{x}). \qquad \text{(B.5)}
\end{aligned}$$

By letting $M_j \to \infty$, it is straightforward to show that (B.5) also holds when $K_{\sigma_j^2}(d\mathrm{y}, \mathrm{x})$ and $\phi(\mathrm{x})$ are general measurable functions. By Wolpert and Ickstadt (1998a), the form of the Laplace exponent in (B.5) implies that (B.4). □

According to the construction of gamma random fields in Theorem 2, in model (3.8),

$$G_j(d\mathrm{x}) = \sum_{m=1}^{\infty} \mu_{j,m} \delta_{\theta_m}(d\mathrm{x}),$$

where $\theta_m \overset{i.i.d.}{\sim} G_0(d\mathrm{x})/G_0(\mathcal{B})$, $\mu_{j,m} = E_1^{-1}\{\zeta_{jm}/G_0(\mathcal{B})\}/\tau$ and $\zeta_{j,m} = \sum_{l=1}^{m} e_{jl}$ with $e_{jl} \overset{i.i.d.}{\sim} Exp(1)$. Based on this representation, we have the following Theorem,

Theorem 7. *Given G_j, σ_j^2, $\mathrm{y}_{j,l}$ and $m_{i,j}$, the conditional distribution of $\mathrm{x}_{j,l}$ is,*

$$Pr(\mathrm{x}_{j,l} = \theta_m \mid G_j, \sigma_j^2, \mathrm{y}_{j,l}, m_{i,j}) \propto \mu_{j,m} k_{\sigma_j^2}(\mathrm{y}_{j,l}, \theta_m). \qquad \text{(B.6)}$$

Proof. By the construction (3.7) of $\mathrm{x}_{j,l}$ in Section 3.2.1 and we can write $G_j(d\mathrm{x}) = \sum_{m=1}^{\infty} \mu_{j,m}\delta_{\theta_m}(d\mathrm{x})$,

$$\pi(\mathrm{x}_{j,l} \mid G_j, \sigma_j^2, \mathbf{y}_{j,l}, m_{i,j}) \propto \sum_{m=1}^{\infty} \mu_{j,m} k_{\sigma_j^2}(\mathrm{y}_{j,l}, \theta_m) I_{\theta_m}(\mathrm{x}_{j,l}),$$

which implies (B.6). □

All other parameters in model (3.8) of the IPGRF can be updated by Metropolis within Gibbs sampling.

APPENDIX C

The Details of Reverse Inference

C.1 Proofs

C.1.1 Proof of Proposition 1

Proof. First, we notice that

$$\begin{aligned}
\pi(\Theta \mid \mathbf{x}_{n+1}, \mathcal{D}_n) &= \frac{\pi(\Theta \mid \mathbf{x}_{n+1}, \mathcal{D}_n)}{\pi(\Theta \mid \mathcal{D}_n)} \pi(\Theta \mid \mathcal{D}_n) \\
&= \frac{\pi(\mathbf{x}_{n+1}, \mathcal{D}_n \mid \Theta)\pi(\Theta)}{\pi(\mathbf{x}_{n+1}, \mathcal{D}_n)} \cdot \frac{\pi(\mathcal{D}_n)}{\pi(\mathcal{D}_n \mid \Theta)\pi(\Theta)} \cdot \pi(\Theta \mid \mathcal{D}_n) \\
&= \frac{\pi(\mathbf{x}_{n+1} \mid \Theta)\pi(\mathcal{D}_n \mid \Theta)\pi(\Theta)}{\pi(\mathbf{x}_{n+1}, \mathcal{D}_n)} \cdot \frac{\pi(\mathcal{D}_n)}{\pi(\mathcal{D}_n \mid \Theta)\pi(\Theta)} \cdot \pi(\Theta \mid \mathcal{D}_n) \\
&= \frac{\pi(\mathbf{x}_{n+1} \mid \Theta)}{\pi(\mathbf{x}_{n+1} \mid \mathcal{D}_n)} \pi(\Theta \mid \mathcal{D}_n) \\
&= \frac{\pi(\mathbf{x}_{n+1} \mid \Theta)\pi(\Theta \mid \mathcal{D}_n)}{\int \pi(\mathbf{x}_{n+1} \mid \Theta)\pi(\Theta \mid \mathcal{D}_n) d\Theta}. \qquad \text{(C.1)}
\end{aligned}$$

Plug (4.14) into (4.1), the probability become

$$\begin{aligned}
&\Pr[T_{n+1} = j \mid \mathbf{x}_{n+1}, \mathcal{D}_n] \\
&\quad = \frac{\int p_j \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j, \Theta)\pi(\Theta \mid \mathcal{D}_n) d\Theta}{\int \sum_{j'=1}^{J} p_{j'} \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j', \Theta)\pi(\Theta \mid \mathcal{D}_n) d\Theta}. \qquad \text{(C.2)}
\end{aligned}$$

C.1.2 Proof of Corollary 1

Proof. In the BHICP model, we assume independence between different types:

$$\begin{aligned}\pi(\mathcal{D}_n \mid \Theta) &= \prod_{j=1}^{J} \pi(\mathbf{x}^j \mid \Theta_j),\\ \pi(\Theta) &= \prod_{j=1}^{J} \pi(\Theta_j).\end{aligned}$$

This implies that

$$\begin{aligned}\pi(\Theta \mid \mathcal{D}_n) &= \frac{\pi(\mathcal{D}_n \mid \Theta)\pi(\Theta)}{\int \pi(\mathcal{D}_n \mid \Theta)\pi(\Theta)d\Theta}\\ &= \frac{\prod_{j=1}^{J}\{\pi(\mathbf{x}^j \mid \Theta_j)\pi(\Theta_j)\}}{\int \cdots \int \prod_{j=1}^{J} \pi(\mathbf{x}^j \mid \Theta_j)\pi(\Theta_j)d\Theta_j}\\ &= \prod_{j=1}^{J} \frac{\pi(\mathbf{x}^j \mid \Theta_j)\pi(\Theta_j)}{\int \pi(\mathbf{x}^j \mid \Theta_j)\pi(\Theta_j)d\Theta_J}\\ &= \prod_{j=1}^{J} \pi(\Theta_j \mid \mathbf{x}^j),\end{aligned}$$

which completes the proof.

C.1.3 Proof of Proposition 2

Proof. The LOOCV posterior predictive probability is

$$\begin{aligned}&\Pr[T_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}]\\ = &\int \Pr[T_i = j \mid \mathbf{x}_i, \Theta]\pi(\Theta \mid \mathbf{x}_i, \mathcal{D}_{-i})d\Theta\\ = &\int \Pr[T_i = j \mid \mathbf{x}_i, \Theta]\frac{\pi(\Theta \mid \mathbf{x}_i, \mathcal{D}_{-i})}{\pi(\Theta \mid \mathbf{x}_i, T_i = t_i, \mathcal{D}_{-i})}\pi(\Theta \mid \mathbf{x}_i, T_i = t_i, \mathcal{D}_{-i})d\Theta.\end{aligned}$$

Note that

$$\begin{aligned}&\frac{\pi(\Theta \mid \mathbf{x}_i, \mathcal{D}_{-i})}{\pi(\Theta \mid \mathbf{x}_i, T_i = t_i, \mathcal{D}_{-i})}\\ = &\frac{\pi(\Theta, \mathbf{x}_i, \mathcal{D}_{-i})\pi(\mathbf{x}_i, T_i = t_i, \mathbf{x}_{-i}, t_{-i})}{\pi(\mathbf{x}_i, \mathcal{D}_{-i}) \cdot \pi(\Theta, \mathbf{x}_i, T_i = t_i, \mathcal{D}_{-i})}\\ = &\frac{\Pr[T_i = t_i \mid \mathbf{x}_i, \mathcal{D}_{-i}]}{\Pr[T_i = t_i \mid \mathbf{x}_i, \Theta]}.\end{aligned}$$

This implies that

$$\begin{aligned}
&\frac{\Pr[t_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}]}{\Pr[T_i = t_i \mid \mathbf{x}_i, \mathcal{D}_{-i}]} \\
&\quad = \int \frac{\Pr[T_i = j \mid \mathbf{x}_i, \Theta]}{\Pr[T_i = t_i \mid \mathbf{x}_i, \Theta]} \pi(\Theta \mid \mathbf{x}_i, T_i = t_i, \mathcal{D}_{-i}) d\Theta \\
&\quad = \int \frac{\pi[\mathbf{x}_i \mid T_i = j, \Theta] p_j}{\pi[\mathbf{x}_i \mid T_i = t_i, \Theta] p_{t_i}} \pi(\Theta \mid \mathbf{x}_i, T_i = t_i, \mathcal{D}_{-i}) d\Theta \\
&\quad := Q_{j,t_i}.
\end{aligned}$$

By $\sum_{j=1}^{J} \Pr[T_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}] = 1$, we have that

$$\frac{1 - \Pr[T_i = t_i \mid \mathbf{x}_i, \mathcal{D}_{-i}]}{\Pr[T_i = t_i \mid \mathbf{x}_i, \mathcal{D}_{-i}]} = \sum_{j \neq t_i} Q_{j,t_i}.$$

This implies

$$\begin{aligned}
\Pr[T_i = t_i \mid \mathbf{x}_i, \mathcal{D}_{-i}] &= \frac{1}{1 + \sum_{j \neq t_i} Q_{j,t_i}}, \\
\Pr[T_i = j \mid \mathbf{x}_i, \mathcal{D}_{-i}] &= \frac{Q_{j,t_i}}{1 + \sum_{j \neq t_i} Q_{j,t_i}}.
\end{aligned}$$

□

C.1.4 Proof of Proposition 3

Proof. The SFIM is computed by

$$\begin{aligned}
&E_{\mathbf{x}_{n+1}|\mathcal{D}_n} \left[\Pr(T_{n+1} = j \mid \mathbf{x}_{n+1} \cup \{\xi\}, \mathcal{D}_n) - \Pr(T_{n+1} = j \mid \mathbf{x}_{n+1}, \mathcal{D}_n) \right] \\
&\quad = E_{\mathbf{x}_{n+1}|\mathcal{D}_n} \big\{ E_{\Theta|\mathcal{D}_n}[\Pr(T_{n+1} = j \mid \mathbf{x}_{n+1} \cup \{\xi\}, \Theta)] \\
&\qquad\qquad - E_{\Theta|\mathcal{D}_n}[\Pr(T_{n+1} = j \mid \mathbf{x}_{n+1}, \Theta)] \big\} \\
&\quad = E_{\mathbf{x}_{n+1},\Theta|\mathcal{D}_n} \left[\Pr(t_{n+1} = j \mid \mathbf{x}_{n+1} \cup \{\xi\}, \Theta) - \Pr(t_{n+1} = j \mid \mathbf{x}_{n+1}, \Theta) \right] \\
&\quad = E_{\mathbf{x}_{n+1},\Theta|\mathcal{D}_n} \left[\frac{p_j \lambda_{1j}(\xi \mid \Theta_j) \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j, \Theta_j)}{\sum_{j'=1}^{T} p_{j'} \lambda_{1j'}(\xi \mid \Theta_{j'}) \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j', \Theta_{j'})} \right] \\
&\qquad - E_{\mathbf{x}_{n+1},\Theta|\mathcal{D}_n} \left[\Pr(T_{n+1} = j \mid \mathbf{x}_{n+1}, \Theta) \right] \\
&= E_{\mathbf{x}_{n+1},\Theta|\mathcal{D}_n} \left[\frac{p_j \lambda_{1j}(\xi \mid \Theta_j) \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j, \Theta_j)}{\sum_{j'=1}^{T} p_{j'} \lambda_{1j'}(\xi \mid \Theta_{j'}) \pi(\mathbf{x}_{n+1} \mid T_{n+1} = j', \Theta_{j'})} \right] - p_j,
\end{aligned}$$

where

$$\begin{aligned}
& E_{\mathbf{x}_{n+1},\Theta|\mathcal{D}_n}\left[\Pr(T_{n+1}=j \mid \mathbf{x}_{n+1},\Theta)\right] \\
= & \int \frac{\pi(\mathbf{x}_{n+1} \mid T_{n+1}=j,\Theta_j)\Pr(T_{n+1}=j)}{\pi[\mathbf{x}_{n+1} \mid \Theta]}\pi(\mathbf{x}_{n+1} \mid \Theta)\pi(\Theta \mid \mathcal{D}_n)d\mathbf{x}_{n+1}d\Theta \\
= & \int \Pr(T_{n+1}=j)\pi(\mathbf{x}_{n+1} \mid T_{n+1}=j,\Theta_j)\pi(\Theta \mid \mathcal{D}_n)d\mathbf{x}_{n+1}d\Theta \\
= & \int \Pr(T_{n+1}=j)\pi(\mathbf{x}_{n+1} \mid T_{n+1}=j,\Theta_j)\pi(\Theta_j \mid \mathcal{D}_n)d\mathbf{x}_{n+1}d\Theta_j \\
= & \Pr(T_{n+1}=j)=p_j.
\end{aligned}$$

This implies that $S_j(\xi) \in [-p_j, 1-p_j]$. $\square$

BIBLIOGRAPHY

BIBLIOGRAPHY

Adolphs, R. (1999), “The human amygdala and emotion,” *Neuroscientist*, 5, 125–137.

Alqallaf, F. and Gustafson, P. (2001), “On cross-validation of Bayesian models,” *Canadian Journal of Statistics*, 29, 333–340.

Antoniak, D. (1974), “Mixtures of Dirichlet Processes With Applications to Bayesian Nonparametric Problems,” *The Annals of Statistics*, 2, 1152–1174.

Baddeley, A. (2010), “Multivariate and Marked Point Processes,” in *Handbook of Spatial Statistics*, eds. Gelfand, A. E., Diggle, P., Guttorp, P., and Fuentes, M., Chapman & Hall/CRC, chap. 21, pp. 371–402.

Baddeley, A., Møller, J., and Waagepetersen, R. (2000), “Non- and semi- parametric estimation of interaction in inhomogeneous point patterns,” *Statistica Neerlandica*, 54, 329–350.

Baddeley, A., Turner, R., Mller, J., and Hazelton, M. (2005), “Residual analysis for spatial point processes (with discussion),” *Journal of the Royal Statistical Society, Series B*, 67, 617–666.

Best, N. G., Ickstadt, K., and Wolpert, R. L. (2000), “Spatial Poisson regression for health and exposure data measured at disparate resolutions,” *Journal of the American Statistical Association*, 95, 1076–1088.

Best, N. G., Ickstadt, K., Wolpert, R. L., Cockings, S., Elliott, P., Bennett, J., Bottle, A., and Reed, S. (2002), “Modeling the Impact of Traffic-Related Air-Pollution On Childhood Respiratory Illness,” in *In Case Studies In Bayesian Statistics, Volume V*, eds. Gatsonis, C., Kass, R., Carlin, B., Carriquiry, A., Gelman., A., Verdinelli, I., and West, M., Springer-Verlag, pp. 183–259.

Blei, D. M., Griffiths, T. L., and Jordan, M. I. (2010), “The nested chinese restaurant process and bayesian nonparametric inference of topic hierarchies,” *Journal of the ACM*, 57, 111–142.

Bondesson, L. (1982), “On simulation from infinitely distributions,” *Advances in Applied Probability*, 855–869.

Boros, G. and Moll, V. (2004), “The Expansion of the Loggamma Function.” in *Irresistible Integrals: Symbolics, Analysis and Experiments in the Evaluation of Integrals*, Cambridge, England: Cambridge University Press, chap. 10.6, pp. 201–203.

Bowman, F. D. (2005), “Spatio-temporal modeling of localized brian activity,” *Biostatistics*, 6, 558–575.

Brix, A. and Møller, J. (2001), “Space-time multitype log Gaussian Cox processes with a view to modeling weed data,” *Scandinavian Journal of Statistics*, 28, 471–488.

Cox, D. R. (1955), “Some statistical models related with series of events,” *Journal of the Royal Statistical Society, Series B*, 17, 129–164.

Damien, P., Laud, P. W., and Smith, A. F. M. (1995), "Approximate random variate generation from infinitely divisible distributions with applications to Bayesian inference," *Journal of the Royal Statistical Society, Series B*, 57, 547–563.

Desikan, R., Sgonne, F., Fischl, B., Quinn, B., Dickerson, B., Blacker, D., Buckner, R., Dale, A., Maguire, R., Hyman, B., Albert, M., and Killiany, R. (2006), "An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest," *Neuroimage*, 31, 968–980.

Diggle, P. (1983), *Statistical analysis of spatial point patterns*, Academic Press.

Diggle, P. J. (1981), "Some graphical methods in the analysis of spatial point patterns," in *Interpreting multivariate data*, ed. Barnett, V., New York: John Wiley & Sons, pp. 55–73.

— (1990), "A point process modeling approach to raised incidence of a rare phenomenon in the vicinity of a prespecified point," *Journal of the Royal Statistical Society, Series A*, 153, 349–362.

Diggle, P. J. and Milne, R. K. (1985), "Bivariate Cox processes: some models for bivariate spatial point patterns," *Journal of the Royal Statistical Society, Series B*, 45, 11–21.

Diggle, P. J. and Rowlingson, B. (1994), "A conditional approach to point process modeling of elevated risk," *Journal of the Royal Statistical Society, Series A*, 157, 433–440.

Diggle, P. J., Zheng, P., and Durr, P. (2005), "Non-parametric estimation of spatial segregation in a multivariate point process," *Applied Statistics*, 54, 645–658.

Eickhoff, S. B., Laird, A. R., Grefkes, C., Wang, L. E., Zilles, K., and Fox, P. T. (2009), "Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty," *Human Brain Mapping*, 30, 2907–2926.

Escobar, M. D. and West, M. (1995), "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, 90, 577–588.

Feng, C., Narayana, S., Lancaster, J. L., Jerabek, P. A., Arnow, T. L., Zhu, F., Tan, L. H., Fox, P. T., and Gao, J. (2004), "CBF changes during brain activation: fMRI vs. PET," *NeuroImage*, 22, 443–446.

Ferguson, T. S. (1973), "A Bayesian Analysis of Some Nonparametric Problems," *The Annals of Statistics*, 1, 209–230.

Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. (2011), "A Sticky HDP-HMM with Application to Speaker Diarization," *The Annals of Applied Statistics*, To appear.

Fox, P. T., Lancaster, J. L., Parsons, L. M., Xiong, J., and Zamarripa, F. (1997), "Functional volumes modeling: theory and preliminary assessment," *Human Brain Mapping*, 5, 306–311.

Friston, K. J., Ashburner, J. T., Kiebel, S. J., Nichols, T. E., and Penny, W. D. (eds.) (2007), *Statistical Parametric Mapping: The Analysis of Functional Brain Images*, Academic Press.

Friston, K. J., Holmes, A. P., Worsley, K. J., Poline, J.-P., Frith, C. D., and Frackowiak, R. S. J. (1994), "Statistical parametric maps in functional imaging: A general linear approach," *Human Brain Mapping*, 2, 189–210.

Gelfand, A. E. (1996), "Inference and monitoring convergence," in *Markov Chain Monte Carlo in Practice*, eds. Gilks, W. R., Richardson, S., and Spiegelhalter, D. J., Chapman& Hall, pp. 131–144.

Gelfand, A. E., Dey, D. K., and Chang, H. (1992), "Model determination using predictive distributions with implementation via sampling-based methods (with discussion)," in *Bayesian Statistics*, eds. Bernardo, J. M., Berger, J. O., Dawid, A. P., and Smith, A. F. M., Oxford University Press, vol. 4, pp. 147–167.

Gelman, A., Garlin, J. B., Stern, H. S., and Rubin, D. B. (2004), *Bayesian Data Analysis*, Chapman & Hall/CRC, 2nd ed.

Gelman, A. and Rubin, D. B. (1992), "Inference from iterative simulation using multiple sequences," *Statistical Science*, 7, 457–472.

Genovese, C. R., Lazar, N. A., and Nichols, T. E. (2002), "Thresholding of statistical maps in functional neuroimaging using the false discovery rate," *Neuroimage*, 15, 870–878.

Gilks, W. R. and Wild, P. (1992), "Adaptive rejection sampling for Gibbs sampling," *Applied Statistics*, 41, 337–348.

Gossl, C., Auer, D., and Fahrmeir, L. (2001), "Bayesian spatiotemporal inference in functional magnetic resonance imaging," *Biometrics*, 57, 554–562.

Hashemi, R. H., Bradley, W. G. J., and Lisanti, C. J. (2004), *MRI: The Basics*, Lippincott Williams & Wilkins, 2nd ed.

Hastie, T., Tibshirani, R., and Friedman, J. (2008), *The elements of statistical learning: data mining, inference, and prediction*, Springer, 2nd ed.

Hoffman, M., Cook, P., and Blei, D. (2008), "Data-driven recomposition using the hierarchical Dirichlet process hidden Markov model," in *Proc. International Computer Music Conference.*

Höogmander, H. and Särkkä, A. (1999), "Multitype spatial point patterns with hierarchical interactions," *Biometrics*, 55, 1051–1058.

Hutchings, M. (1979), "Standing crop and pattern in pure stands of Mercurialis perennis and Rubus fruticosus in mixed deciduous woodland," *Oikos*, 31, 351–357.

Ickstadt, K. and Wolpert, R. (1999), "Spatial Regression for Marked Point Processes (with discussion)," in *Bayesian Statistics*, eds. Bernardo, J., Berger, J., Dawid, A., and Smith, A., Oxford: Oxford University Press, chap. 6, pp. 323–341.

Illian, J., Penttinen, A., Stoyan, H., and Stoyan, D. (2008), *Statistical Analysis and Modelling of Spatial Point Patterns*, John Wiley & Sons.

Illian, J. B., Møller, J., and Waagepetersen, R. P. (2009), "Hierarchical spatial point process analysis for a plant community with high biodiversity," *Environmental and Ecological Statistics*, 16, 389–405.

Jezzard, P., Matthews, P. M., and Smith, S. M. (2001), *Functional MRI: An Introduction to Methods*, Oxford University Press.

Kang, J., Johnson, T. D., Nichols, T. E., and Wager, T. D. (2011), "Meta analysis of functional neuroimaging data via Bayesian spatial point processes," *Journal of the American Statistical Association*, 106, 124–134.

Katanoda, K., Matsuda, Y., and Sugishita, M. (2002), "A spatio-temporal regression model for the analysis of functional MRI data," *NeuroImage*, 17, 1415–1428.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K., and Wager, T. D. (2008), "Functional grouping and corticalsubcortical interactions in emotion: A meta-analysis of neuroimaging studies," *NeuroImage*, 42, 998–1031.

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., Turner, R., Cheng, H., Brady, T. J., and Rosen, B. R. (1992), "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields," *Proceedings of the National Academy of Sciences*, 89, 5675–5679.

Lawson, A. B., Biggeri, A. B., Boehning, D., Lesaffre, E., Viel, J.-F., Clark, A., Schlattmann, P., and Divino, F. (2000), "Disease mapping models: an empirical evaluation," *Statistics in Medicine*, 19, 2217–2241.

Lazar, N. A. (2008), *The Statistical Analysis of Functional MRI Data*, Springer.

Lim, C. Y. and Dass, S. C. (2011), "Assessing Fingerprint Individuality Using EPIC: A Case Study in the Analysis of Spatially Dependent Marked Processes," *Technometrics*, 53, 112–124.

Lindquist, M. A. (2008), "The Statistical Analysis of fMRI Data," *Statistical Science*, 23, 439–464.

Lotwick, H. W. and Silverman, B. W. (1982), "Methods for analysing spatial point processes of several types of points," *Journal of the Royal Statistical Society, Series B*, 44, 406–413.

Luo, W. L. and Nichols, T. E. (2003), "Diagnosis and Exploration of Massively Univariate fMRI Models," *NeuroImage*, 19, 1014–1032.

MacEachern, S. N. (1999), "Dependent Nonparametric Processes," in *Proceedings of the Bayesian Statistical Science Section*, American Statistical Association.

Mardia, K. V., Kent, J. T., and Bibby, J. M. (1979), *Multivariate Analysis*, Academic Press.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., Zilles, K., Woods, R., Paus, T., Simpson, G., Pike, B., Holmes, C., Collins, L., Thompson, P., MacDonald, D., Iacoboni, M., Schormann, T., Amunts, K., Palomero-Gallagher, N., Geyer, S., Parsons, L., Narr, K., Kabani, N., Le Goualher, G., Boomsma, D., Cannon, T., Kawashima, R., and Mazoyer, B. (2001), "A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM)." *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 356, 1293–1322.

Møller, J., Syversveen, A. R., and Waagepetersen, R. P. (1998), "Log Gaussian Cox processes," *Scandinavian Journal of Statistics*, 25, 451–482.

Møller, J. and Waagepetersen, R. (2007), "Modern statistics for spatial point processes (with discussion)," *Scandinavian Journal of Statistics*, 34, 643–711.

Møller, J. and Waagepetersen, R. P. (2004), *Statistical Inference and Simulation for Spatial Point Processes*, Chapman and Hall/CRC.

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

Nielsen, F. A. and Hansen, L. K. (2002), "Modeling of activation data in the BrainMap database: detection of outliers," *Human Brain Mapping*, 15, 146–156.

Niemi, A. and Fernández, C. (2010), "Bayesian Spatial Point Process Modeling of Line Transect Data," *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 327–345.

Ogawa, S. and Lee, T. M. (1992), "Functional brain imagining with physiologically sensitive image signals," *Journal of Magnetic Resonance Imaging*, 2(P)-WIP(Suppl), S22.

Ogawa, S., Lee, T. M., Nayak, A. S., and Glynn, P. (1990), "Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields," *Mangnetic Resonance Medicine*, 14, 68–78.

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

Pereira, F., Mitchell, T., and Botvinick, M. (2009), “Machine Learning Classifiers and fMRI: A Tutorial Overview,” *NeuroImage*, 45, S199–S209.

Phelps, E. A. and LeDoux, J. E. (2005), “Contributions of the amygdala to emotion processing: from animal models to human behavior,” *Neuron*, 48, 175–187.

Poldrack, R. (2006), “Can cognitive processes be inferred from neuroimaging data?” *Trends in Cognitive Sciences*, 10, 59–63.

Poldrack, R., Halchenko, Y., and Hanson, S. (2009), “Decoding the large-scale structure of brain function by classifying mental states across individuals,” *Psychological Science*, 20, 1364–1372.

Press, S. J. (1982), *Applied Multivariate Analysis*, Dover Publications, 2nd ed.

Preston, C. J. (1975), “Spatial birth-and-death processes,” *Bulletin of the International Statistical Institute*, 46, 371–391.

Radua, J. and Mataix-Cols, D. (2009), “Voxel-wise meta-analysis of grey matter changes in obsessivecompulsive disorder,” *The British Journal of Psychiatry*, 195, 393–402.

Raichle, M. (2003), “Functional brain imaging and human brain function,” *The Journal of Neuroscience*, 23, 3959–3962.

Raichle, M. E. (2006), “Functional Neuroimaging: A Historical and Physiological Perspective,” in *Handbook of Functional Neuroimaging of Cognition, 2nd Edition*, eds. Cabeza, R. and Kingstone, A., Cambridge (MA): The MIT Press, chap. 1, pp. 3–20.

Rodriguez, A., Dunson, D. B., and Gelfand, A. E. (2008), “The Nested Dirichlet Processes,” *Journal of the American Statistical Association*, 103, 1131–1154.

Salimi-Khorshidi, G., Smith, S. M., Keltner, J. R., Wager, T. D., and Nichols, T. E. (2009), “Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies,” *NeuroImage*, 45, 810–823.

Springer, C., Patlak, C., Palyka, I., and Huang, W. (1999), “Principles of susceptibility contrast-based functional MRI: The sign of the functional MRI response,” in *Functional MRI*, eds. Moonen, C. T. W. and Bandettini, P. A., Springer, chap. 9, pp. 90–102.

Stoyan, D. and Penttinen, A. (2000), “Recent Applications of Point Process Methods in Forestry Statistics,” *Statistical Science*, 15, 61–78.

Strother, S. C. (2006), “Evaluating fMRI preprocessing pipelines,” *Engineering in Medicine and Biology Magazine, IEEE*, 25, 27–41.

Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. (2006), “Hierarchical Dirichlet processes,” *Journal of the American Statistical Association*, 101, 1566–1581.

Thirion, B., Pinel, P., Mriaux, S., Roche, A., Dehaene, S., and Poline, J.-B. (2007), “Analysis of a large fMRI cohort: Statistical and methodological issues for group analyses,” *Neuroimage*, 35, 105–120.

Turkeltaub, P. E., Eden, G. F., Jones, K. M., and Zeffiro, T. A. (2002), “Meta-analysis of the functional neuroanatomy of single-word reading: method and validation,” *NeuroImage*, 16, 765–780.

Tziortzi, A. C., Searle, G. E., Tzimopoulou, S., Salinas, C., Beaver, J. D., Jenkinson, M., Laruelle, M., Rabiner, E. A., and Gunn, R. N. (2010), "Imaging dopamine receptors in humans with [(11)C]-(+)-PHNO: Dissection of D3 signal and anatomy." *NeuroImage*.

van Lieshout, M. and Baddeley, A. J. (2001), "Extrapolating and interpolating spatial patterns," in *In Spatial cluster modelling, A.B. Lawson and D.G.T. Denison (Eds.) Boca Raton: Chapman and Hall/CRC*, Press, pp. 61–86.

van Lieshout, M. N. M. and Baddeley, A. J. (1999), "Standing crop and pattern in pure stands of Mercurialis perennis and Rubus fruticosus in mixed deciduous woodland," *Scandinavian Journal of Statistics*, 26, 511–532.

Vehtari, A. and Lampinen, J. (2002), "Bayesian Model Assessment and Comparison Using Cross-Validation Predictive Densities," *Neural Computation*, 14, 2439–2468.

Wager, T. D., Hernandez, L., Jonides, J., and Lindquist, M. (2007a), "Elements of functional neuroimaging," in *Handbook of Psychophysiology*, eds. Cacioppo, J. T., Tassinary, L. G., and Berntson, G. G., Cambridge University Press, 4th ed., pp. 19–55.

Wager, T. D., Jonides, J., and Reading, S. (2004), "Neuroimaging studies of shifting attention: a meta-analysis," *NeuroImage*, 22, 1679–1693.

Wager, T. D., Lindquist, M., and Kaplan, L. (2007b), "Meta-analysis of functional neuroimaging data: current and future directions," *Social Cognitive and Affective Neuroscience*, 2, 150–158.

Wager, T. D., Lindquist, M. A., Nichols, T. E., Kober, H., and van Snellenberg, J. X. (2009), "Evaluating the consistency and specificity of neuroimaging data using meta-analysis," *NeuroImage*, 45, 210–221.

Wills, A. P. and Hector, L. G. (1924), "The Magnetic Susceptibility of Oxygen, Hydrogen and Helium," *Physical Review*, 23, 209–220.

Wolpert, R. L. and Ickstadt, K. (1998a), "Poisson/gamma random field models for spatial statistics," *Biometrika*, 85, 251–267.

— (1998b), "Simulation of Lévy random fields," in *Practical Nonparametric and Semiparametric Bayesian Statistics*, eds. Dey, D., Müller, P., and Sinha, D., New York: Springer Verlag, pp. 227–242.

Woodard, D. B., Wolpert, R. L., and OConnell, M. A. (2010), "Spatial Inference of Nitrate Concentrations in Groundwater," *Journal of Agricultural, Biological, and Environmental Statistics*, 15, 209–227.

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

Xing, E. P. and Sohn, K. A. (2007), "Hidden Markov Dirichlet process: Modeling genetic inference in open ancestral space," *Bayesian Analysis*, 2, 501–528.

Yadrenko, M. I. (1983), *Spectral theory of random fields*, Optimization Software.

Yarkoni, T. (2009), "Big correlations in little studies: Inflated fMRI correlations reflect low statistical powerCommentary on Vul et al. (2009)," *Perspective on Psychological Science*, 4, 294–298.

Yarkoni, T., Poldrack, R. A., Nichols, T. E., Van Essen, D. C., and Wager, T. D. (2011), "Large-scale lexical decoding of human brain activity," *Unpublished Manuscript.*

Yarkoni, T., Poldrack, R. A., Van Essen, D. C., and Wager, T. D. (2010), "Cognitive neuroscience 2.0: building a cumulative science of human brain function," *Trends in Cognitive Sciences*, 14, 489–496.