Research Article

Detecting Spatial Clusters via a Mixture of Dirichlet Processes

Meredith A. Ray ®,[1] Jian Kang,[2] and Hongmei Zhang ®[3]

[1]*Division of Epidemiology, Biostatistics, and Environmental Health, University of Memphis, 220 Robison Hall,*
 Memphis, TN 38152, USA
[2]*Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109, USA*
[3]*Division of Epidemiology, Biostatistics, and Environmental Health, University of Memphis, 224 Robison Hall,*
 Memphis, TN 38152, USA

Correspondence should be addressed to Meredith A. Ray; maray@memphis.edu

We proposed an approach that has the ability to detect spatial clusters with skewed or irregular distributions. A mixture of Dirichlet processes (DP) was used to describe spatial distribution patterns. The effects of different batches of data collection efforts were also modeled with a Dirichlet process. To cluster spatial foci, a birth-death process was applied due to its advantage of easier jumping between different numbers of clusters. Inferences of parameters including clustering were drawn under a Bayesian framework. Simulations were used to demonstrate and assess the method. We applied the method to an fMRI meta-analysis dataset to identify clusters of foci corresponding to different emotions.

1. Introduction

This work was motivated by a study aiming to detect centers of activated foci from a function magnetic resonance imaging (fMRI) metadataset. In summary, fMRI metadata is a collection of fMRI studies identifying areas of the brain that are significantly activated by stimuli to examine a specific outcome. FMRIs are expensive which leads to small sample sizes and therefore can use metadata to increase sample size and power. To identify spatial clusters, finite mixture models are generally implemented [1–3]. Mixture components, representing a cluster, typically share a common parametric family with each component containing respective parameters [1, 2]. Each component also has a mixing proportion or weight that is respective to the frequency of data in that component [1]. Because of the model's ease of implementation, this allows various applications such as pattern recognition, computer vision, signal and image analysis, and machine learning to list a few [4].

One commonly used distribution in the aforementioned finite mixtures is the normal distribution [1, 5], appreciating its established properties and, in the Bayesian context, conjugacy. However, when it comes to clustering, assuming normality for each part of the mixture can potentially lead to oversensibility, e.g., when one cluster is formed by a mixture of two normal distributions but with rather close centers. This type of oversensibility in many research fields should be avoided; one example is emotional foci inferred from brain imaging data, where a certain emotion is covered by a wide region.

To infer the number of clusters, under the Bayesian framework, different methods have been proposed. Reversible jump Markov chain Monte Carlo has been commonly used to infer the number of clusters [6, 7]. At each iteration, a decision is made between splitting one cluster to two, combining two clusters to one, or no movement. One potential difficulty of this approach is the risk of being trapped at a local maximum. Recently, the Dirichlet process (DP) has been used often to estimate the number of clusters [8, 9]. This process has the ability to capture irregular patterns. DP has the ability to detect clusters without being burdened about additional clustering parameters. However, this feature of clustering also has an inherit weakness in that it tends to produce more clusters, making interpretations more difficult.

To overcome the aforementioned gaps, we implemented a mixture of Dirichlet processes (DP). These processes have the ability to describe irregular patterns [8, 9] and by using it as our common parametric family allows the model to

identify more complex patterns than the normal distribution would be able to identify. Furthermore, motivated from our previous work with the spatial Cox process application in [10], we elected to incorporate the birth-death process to statistically determine the number of clusters. Compared to other clustering approaches noted earlier, the birth-death process has the advantage of quick convergence and, by controlling birth rate, embraces a potential of generating redundant clusters.

The remainder of the article is organized as follows. In Section 2, we introduce the model structure, notation, and priors and hyperpriors, simulations are presented in Section 3, the application of the model to an fMRI meta-analysis dataset is in Section 4, and conclusions and discussion are in Section 5.

2. Methods

2.1. The Model. We let s_{rj} denote the (x, y, z) coordinate of a spatial point in a three-dimensional space, in particular, point j in group (study) r, $j = 1, \ldots, J_r$ and $r = 1, \ldots, R$. We have $\sum_{r=1}^{R} J_r = n$, where n is the total number of observed points. It straightforwardly follows that $\boldsymbol{s} = (s_{1,1} \cdots s_{R,J_R})$ represents all points in the study. We model s_{rj} as

$$s_{rj} = p_r + \theta_{rj} + \epsilon, \tag{1}$$

where p_r denotes the effect of group r, while θ_{rj} represents the mean of s_{rj} for the jth point in group r after adjusting for group effects, p_r, and ϵ denotes some random error. By modeling the random error as a standard multivariate normal distribution, the distribution of s_{rj} satisfies

$$s_{rj} \sim MVN_3 \left(p_r + \theta_{rj}, \Sigma \right) \tag{2}$$

with

$$P \left(s_{rj} \mid p_r, \theta_{rj}, \Sigma \right) = (2\pi)^{-3/2} \, |\Sigma|^{-1/2}$$
$$\cdot \exp \left[-\frac{1}{2} \left(s_{rj} - p_r - \theta_{rj} \right)' \Sigma^{-1} \left(\left(s_{rj} - p_r - \theta_{rj} \right) \right) \right], \tag{3}$$

where $\Sigma = \sigma^2 I_3$ is the covariance matrix.

2.2. Prior and Hyperprior Distributions. We start from the prior distribution of θ_{rj}. To detect underlying clusters of s_{rj} due to similarities of θ_{rj}, we describe the prior of θ_{rj} as a mixture of distribution G_k, $k = 1, \ldots, K$. Common choices of G_k are normal distributions. To improve flexibility, we relax such normalizing assumptions in the mixture and assume G_k is generated from a DP, i.e., $G_k \sim DP(\alpha, G_{0k})$, where α is the precision parameter and G_{0k} is a base distribution and taken as $G_{0k} = MVN_3(\boldsymbol{\mu_k}, \Sigma_k)$. In particular, $\theta_{rj} \sim \sum_{k=1}^{K} \pi_k G_k$, where $0 < \pi_k < 1$ such that $\sum_{k=1}^{K} \pi_k = 1$. For the number of clusters K, we assign a truncated Poisson prior distribution to K, $P(K) = (\lambda^K / K!) \exp(-\lambda)$, $K = 1, \ldots, n$. We assign a Dirichlet distribution with parameter 1 to $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$, which implies that $\boldsymbol{\pi}$ is k-dimensional uniformly distributed.

The prior distribution of $\boldsymbol{\mu_k}$ from the base distribution G_{0k} is chosen to be $\boldsymbol{\mu_k} \sim MVN_3(\boldsymbol{\xi}, \kappa^{-1})$, where $\boldsymbol{\xi} = (\xi_1, \xi_2, \xi_3)$, which are known and set as the midpoint of observed interval of variation of the data. Parameter κ is set as

$$\kappa = \begin{bmatrix} \dfrac{1}{R_1^2} & 0 & 0 \\ 0 & \dfrac{1}{R_2^2} & 0 \\ 0 & 0 & \dfrac{1}{R_3^2} \end{bmatrix}, \tag{4}$$

where R_1^2, R_2^2, and R_3^2 are the range of the data for each dimension. This prior setting is adopted from [1] and we feel it is reasonable for this setting given the fact that the number and location of clusters are unknown. We let $\Sigma_k = \sigma_k I_3$ with $\sigma_k \sim IG(20, 0.5)$ since the range of the observed data is small. For group effect, p_r, we assume it is small with $p_r \sim G_p$, where $G_p \sim DP(\alpha_p, G_{0p})$ with α_p specified later and $G_{0p} = TMVN_3(\boldsymbol{\mu_p}, \Sigma_p, lower = -l, upper = l)$, where the lower and upper limits are defined as 10% of the absolute range of the data. We let $\boldsymbol{\mu_p} \sim MVN_3(\boldsymbol{0}, I_3)$ and $\Sigma_p = \sigma_p^2 I_3$, with $\sigma_p^2 \sim IG(5, 0.5)$. The variance component of the random error, $\Sigma = \sigma^2 I_3$, σ^2 is assumed to follow a relatively noninformative Inverse Gamma (IG) distribution, $\sigma^2 \sim IG(0.5, 0.5)$. The precision parameters α and α_p are selected by minimizing the deviance information criterion (DIC) [11, 12].

2.3. Conditional Posterior Distributions and Posterior Computing. Sampling parameter estimates from their posterior distributions can be achieved via Gibbs sampler, in which the statistical inference on the number of clusters is modeled using the birth-death process. The birth-death process is one type of continuous-time Markov chain originally introduced in [13]. This process is often used to simulate realizations of point processes as they can be difficult to directly sample from [1]. These realizations are further used for likelihood inferences for model parameters [1]. The birth-death scheme allows events to randomly occur throughout the chain; these events are either a "birth" or "death." If a birth occurs, the number of components increases by one, while if a death occurs, the number of components decreases by one.

Recall considering a finite mixture prior for θ_{rj} such that all θ_{rj} are assumed independently distributed with each generated from one of K Dirichlet processes denoted as G_k, i.e.,

$$P \left(\theta_{rj} \mid K, \boldsymbol{\pi}, \boldsymbol{\phi}, \alpha \right) = \sum_{k=1}^{K} \pi_k G_k \left(\alpha, G_{0k} \right), \tag{5}$$

where each $G_k(\alpha, G_{0k})$ represents a DP but K is unknown, $\boldsymbol{\pi} = (\pi_1, \ldots, \pi_K)$ are the mixing proportions, and $\boldsymbol{\phi} = (\boldsymbol{\mu_1}, \Sigma_k, \ldots, \boldsymbol{\mu_K}, \Sigma_K)$ are the component specific parameters for each DP. For cluster assignment, we introduce an index variable Z_{rj} that indicates the assignment of observation rj and Z_{rj} takes the values of 1 to K. Denoted by $z_{rj} \in \boldsymbol{z}$, where $\boldsymbol{z} = (z_{1,1}, \ldots, z_{R,J_R})$ represents the realization of independently and identically distributed discrete random

variables $\mathbf{Z} = (Z_{1,1}, \ldots, Z_{R,J_R})$ with probability mass function,

$$P\left(Z_{rj} = k \mid \boldsymbol{\pi}, \boldsymbol{\phi}, \eta\right) = \pi_k \quad (k = 1, \ldots, K). \qquad (6)$$

The joint posterior distribution is proportional to

$$
\begin{aligned}
P\left(\boldsymbol{\vartheta} \mid \boldsymbol{s}\right) &\propto \prod_r^R P\left(p_r \mid G_p\right) P\left(G_p \mid \alpha, G_{0p}\right) \\
&\quad \cdot G_{0p}\left(p_r; \mu_p, \sigma_p^2\right) P\left(\mu_p\right) P\left(\sigma_p^2\right) \\
&\quad \cdot \prod_j^{J_r} P\left(s_{rj} \mid \boldsymbol{\phi}\right) \\
&\quad \times P\left(\theta_{rj} \mid \boldsymbol{z}, \boldsymbol{\pi}, G_1, \ldots, G_K\right) \\
&\quad \cdot \prod_{k=1}^K P\left(G_k \mid \alpha, G_{0k}\right) G_{0k}\left(\theta_{rj}; \mu_k, \sigma_k^2\right) \\
&\quad \cdot P\left(\mu_k\right) P\left(\sigma_k^2\right) \times P\left(\boldsymbol{\pi}\right) P\left(K\right) P\left(\mathbf{Z}\right) P\left(\sigma^2\right),
\end{aligned}
\qquad (7)
$$

where $\boldsymbol{\vartheta} = (p_r, \theta_{rj}, \sigma^2, \alpha_p, \boldsymbol{\mu_p}, \sigma_p^2, K, \boldsymbol{\mu_1}, \ldots, \boldsymbol{\mu_K}, \sigma_1^2, \ldots, \sigma_K^2, \alpha, \boldsymbol{\pi})$ is a vector of all estimable parameters. From here, the birth-death algorithm and Markov chain can be described for introducing and assigning clusters for θ_{rj}:

(1) Starting with the initial model $y = \{(\pi_1, \phi_1), \ldots, (\pi_K, \phi_K)\}$, let (π_k, ϕ_k) represent the mixing proportion and cluster specific parameters for the unique θ_{rj} clusters. Let the birth rate be $\beta(y) = \lambda_b$.

(2) Calculate the death rate for each component:

$$\delta_k\left(y\right) = \frac{L\left(y \setminus (\pi_k, \phi_k)\right)}{L\left(y\right)} \frac{P\left(K - 1 \mid \alpha, \cdot\right)}{K P\left(K \mid \alpha, \cdot\right)} \qquad (8)$$

$$(k = 1, \ldots, K).$$

(3) Calculate the total death rate $\delta(y) = \sum_{k=1}^K \delta_k(y)$. To quicken convergence, we elected not to model the time to next jump as exponential and allowed an event to occur at each iteration of the Markov chain.

(4) Simulate the type of event, birth or death, with the respective probabilities:

$$P\left(birth\right) = \frac{\beta\left(y\right)}{\beta\left(y\right) + \delta\left(y\right)},$$

$$P\left(death\right) = \frac{\delta\left(y\right)}{\beta\left(y\right) + \delta\left(y\right)}. \qquad (9)$$

(5) Adjust the model y to reflect the birth or death by the following:

    (i) Birth: Simulate a new component (π_{K+1}, ϕ_{K+1}) from each parameter's respective (independent) prior distributions, π_{K+1} from $K(1 - \pi)^{K-1}$ and ϕ_{K+1} from base distribution, such that the

model becomes $y = \{(\pi_1, \phi_1), \ldots, (\pi_K, \phi_K), (\pi_{K+1}, \phi_{K+1})\}$. It can be mentioned that $K(1 - \pi)^{K-1}$ is the Beta distribution with parameters $(1, K)$ and can be easily simulated from $Y_1 \sim G(1, 1)$ and $Y_2 \sim G(K, 1)$ such that $\pi_{K+1} = Y_1/(Y_1 + Y_2)$.

    (ii) Death: Select a component to die with the probabilities $\delta_k(y)/\delta(y)$ for $k = 1, \ldots, K$ such that the model becomes $y = \{(\pi_1, \phi_1), \ldots, (\pi_{K-1}, \phi_{K-1})\}$.

(6) Given the current state of the model at time t, simulate values for all remaining parameters.

(7) Go to step (2).

Incorporating the birth-death process into our model, we need to further define the likelihood for removing cluster i:

$$
\begin{aligned}
P\left(\boldsymbol{\Phi} \setminus (\pi_i, G_i) \mid \boldsymbol{s}\right) &\propto \prod_r^R P\left(p_r \mid G_p\right) P\left(G_p \mid \alpha, G_{0p}\right) \\
&\quad \cdot G_{0p}\left(p_r; \mu_p, \sigma_p^2\right) P\left(\mu_p\right) P\left(\sigma_p^2\right) \prod_j^{J_r} P\left(s_{rj} \mid \boldsymbol{\phi}\right) \\
&\quad \times P\left(\theta_{rj} \mid \boldsymbol{z}, \boldsymbol{\pi}, G_1, \ldots, G_K\right) \prod_{k=1}^{K^{(i)}} P\left(G_k \mid \alpha, G_{0k}\right) \\
&\quad \cdot G_{0k}\left(\theta_{rj}; \mu_k, \sigma_k^2\right) \times P\left(\mu_k\right) P\left(\sigma_k^2\right) P\left(\boldsymbol{\pi}\right) P\left(K\right) P\left(\mathbf{Z}\right) \\
&\quad \cdot P\left(\sigma^2\right),
\end{aligned}
\qquad (10)
$$

where $K^{(i)} = i \notin (1, \ldots, K)$. The birth-death process is conditional on the prespecified birth rate, λ_b. By setting this birth rate, which controls how often a "birth" of a new component occurs, equal to λ as suggested and done in [1], this computationally allows the death rates to be a likelihood ratio absent of λ. In other words, the likelihood of the data drives the death rates and ultimately the decision of a new cluster. Given that decision is a "birth," the new cluster's parameters $\boldsymbol{\mu_{K+1}}, \sigma_{K+1}^2$, and π_{K+1} are sampled from their prior distributions:

$$
\begin{aligned}
\boldsymbol{\mu_{K+1}} &\sim MVN_3\left(\boldsymbol{\xi}, \kappa^{-1}\right) \\
\sigma_{K+1}^2 &\sim IG\left(20, 0.5\right) \qquad (11) \\
\pi_{K+1} &\sim K\left(1 - \boldsymbol{\pi}\right)^{K-1}.
\end{aligned}
$$

The mixing proportions are adjusted by multiplying all current proportions by $(1 - \pi_{K+1})$ if a birth occurs or dividing by $(1 - \pi_i)$ if a death occurs.

To simulate values for all remaining parameters and hyperparameters, we implement the Gibbs sampler. Conditional posterior distributions are listed below. Note that "·" denotes data and other parameters not listed. The conditional posterior of Z_{rj} is

$$
\begin{aligned}
P&\left(Z_{rj} = k \mid \boldsymbol{\pi}, \theta_{rj}, G_k, \cdot\right) \\
&\propto P\left(\theta_{rj} \mid Z_{rj} = k, G_k, \cdot\right) P\left(G_k \mid \cdot\right) P\left(Z_{rj} = k\right)
\end{aligned}
$$

$$= \left\{ \frac{\alpha MVN_3 \left(\boldsymbol{\mu_k}, \Sigma_k \right)}{\alpha + n_k - 1} + \frac{\sum_{c=1}^{C_k} \delta_c \left(\theta_{rj} \right)}{\alpha + n_k - 1} \right\} \pi_k,$$

(12)

where $c = 1, \ldots, C_k$ is the number of subclusters for cluster $k \in (1, \ldots, K)$, n_k is the number of foci in cluster k, and $\delta_c(\theta_{rj})$ denotes the unit point mass,

$$\boldsymbol{\pi} \mid \mathbf{Z} \sim Dir \left(n_1 + 1, \ldots, n_K + 1 \right),$$

(13)

and

$$P \left(\theta_{rj} \mid k, \cdot \right) \propto \prod_{Z_{rj} \in k} P \left(s_{rj} \mid p_r, \theta_{rj}, \cdot \right)$$

$$\cdot P \left(\theta_{rj} \mid Z_{rj} = k, G_k \right) P \left(G_k \mid G_{0k}, \alpha \right)$$

$$\cdot G_{0k} \left(\theta_{rj}; \boldsymbol{\mu_k}, \sigma_k^2 \right) = \prod_{Z_{rj} \in k} \left\{ MVN_3 \left(p_r + \theta_{rj}, \Sigma \right) \right\} \quad (14)$$

$$\cdot \left\{ \frac{\alpha MVN_3 \left(\boldsymbol{\mu_k}, \Sigma_k \right)}{\alpha + n_k - 1} + \frac{\sum_{q=1, q \neq rj}^{n_k} \delta_{\theta_q} \left(\theta_{rj} \right)}{\alpha + n_k - 1} \right\},$$

which is the distribution of a DP with $\delta_{\theta_q}(\theta_{rj})$ being the unit point mass and n_k the number of foci in some cluster $k \in (1, \ldots, K)$,

$$\boldsymbol{\mu_k} \mid \cdot \sim MVN_3 \left(\left(\kappa + n_k \sigma_k^{-2} I_3 \right)^{-1} \right.$$

$$\left. \cdot \left(\kappa \boldsymbol{\xi} + n_k \sigma_k^{-2} I_3 \overline{\theta}_{rj} \right), \left(\kappa + n_k \sigma_k^{-2} I_3 \right)^{-1} \right),$$

(15)

where $\overline{\theta}_{rj} = \sum_{rj \in k} \theta_{rj} / n_k$ and denotes the average of all θ_{rj} in cluster k,

$$\sigma_k^2 \mid \cdot \sim IG \left(\frac{3 \times n_k}{2} + 20, 0.5 \right.$$

$$\left. + \frac{1}{2} \sum_{Z_{rj} \in k} \left(\theta_{rj} - \boldsymbol{\mu_k} \right)' \left(\theta_{rj} - \boldsymbol{\mu_k} \right) \right),$$

(16)

where n_k is the number of foci in cluster k, $k = 1, \ldots, K$. The conditional posterior of p_r is generated from a DP again:

$$P \left(p_r \mid \cdot \right) \propto \prod_{r=1}^{R} P \left(s_{rj} \mid p_r, \theta_{rj}, \cdot \right) P \left(\theta_{rj} \mid \cdot \right) P \left(p_r \mid G_p \right)$$

$$\cdot P \left(G_p \mid G_{0p}, \alpha_p \right) G_{0p} \left(p_r; \mu_p, \sigma_p^2 \right)$$

$$= \prod_{r=1}^{R} \left\{ MVN_3 \left(p_r + \theta_{rj}, \Sigma \right) \right\}$$

(17)

$$\cdot \left\{ \frac{\alpha_p TMVN_3 \left(\boldsymbol{\mu_p}, \Sigma_p, -l, l \right)}{\alpha_p + R - 1} \right.$$

$$\left. + \frac{\sum_{q=1, q \neq r}^{R} \delta_{p_q} \left(p_r \right)}{\alpha_p + R - 1} \right\},$$

where $\delta_{p_q}(p_r)$ is the unit point mass. The conditional posterior distributions for the related hyperparameters are

$$\boldsymbol{\mu_p} \mid \cdot \sim MVN_3 \left(\left(1 + R\sigma_p^{-2} \right)^{-1} \right.$$

$$\left. \cdot \left(R\sigma_p^{-2} \left(\overline{p_r} \right) \right), \left(1 + R\sigma_p^{-2} \right)^{-1} I_3 \right),$$

(18)

where $\overline{p_r} = \sum_{r \in R} p_r / R$ and represents the average group effect and

$$\sigma_p^2 \mid \cdot \sim IG \left(\frac{3 \times R}{2} + 5, \frac{\sum_{r=1}^{R} \left(p_r - \boldsymbol{\mu_p} \right)^2 + 1}{2} \right).$$

(19)

Lastly, the sampling distribution for Σ is

$$\sigma^2 \mid \cdot \sim IG \left(\frac{3 \times n + 1}{2}, \right.$$

$$\left. \frac{1 + \sum_{Z_{rj} \in n} \left(s_{rj} - p_r - \theta_{rj} \right)' \left(s_{rj} - p_r - \theta_{rj} \right)}{2} \right).$$

(20)

The sampling of unique values for θ_{rj} and p_r can be performed using Neal's algorithm 8 [14]. It works by introducing m auxiliary parameters that are independent to other parameters to represent potential values for θ_{rj} and p_r [14]. Algorithm 8 for updating clustering assignments, denoted as c, is as follows:

(i) The state of the Markov chain consists of $c = \{c_1, \ldots, c_n\}$ and $\Phi = (\phi_c; c \in c_1, \ldots, c_n)$ with ϕ_c denoting cluster parameters, e.g., θ_c in our application. Repeatedly sample as follows:

(ii) For $i = 1, \ldots, n$, let k^- be the number of distinct c_l for $l \neq i$, and $h = k^- + m$. Label these c_l with values in $\{1, \ldots, k^-\}$. If $c_i = c_l$ for some $l \neq i$, draw values independently from base distribution G_0 for those ϕ_c for which $k^- < c \leq h$. If $c_i \neq c_l$ for all $l \neq i$, let c_i have the label $k^- + 1$, and draw values independently from G_0 for those ϕ_c for which $k^- + 1 < c \leq h$. Draw a new value for c_i from $\{1, \ldots, h\}$ using the following probabilities:

$$P \left(c_i = c \mid c_{-i}, y_i, \phi_1, \ldots, \phi_h \right)$$

$$= \begin{cases} \dfrac{n_{-i,c}}{n - 1 + \alpha} F \left(y_i, \phi_c \right) & \text{for } 1 \leq c \leq k^- \\ \dfrac{(\alpha/m)}{n - 1 + \alpha} F \left(y_i, \phi_c \right) & \text{for } k^- < c \leq h, \end{cases} \quad (21)$$

where $F(y_i, \theta_c)$ is the likelihood with θ_c and observation i, y_i, involved. In our case, it is s_{ij},

(iii) where $n_{-i,c}$ is the number of c_l for $l \neq i$ that are equal to c and b is the appropriate normalizing constant. Change the state to contain only those ϕ_c that are now associated with one or more observation.

(iv) For all $c \in \{c_1, \ldots, c_n\}$, draw new values from $\phi_c \mid y_i$ such that $c_i = c$, or perform some other update to ϕ_c that leaves this distribution invariant [14].

Thus, the value for θ_{rj} for those foci in cluster k and subcluster c may be sampled from

$$\theta_{k,c} \mid \cdot \sim MVN_3 \left(\left(\sigma_k^{-2} + n_{k,c}\sigma^{-2} \right)^{-1} \right.$$
$$\left. \cdot \left(\sigma_k^{-2}\boldsymbol{\mu_k} + n_{k,c}\sigma^{-2}\overline{(s_{rj} - p_r)} \right), \left(\sigma_k^{-2} + n_k\sigma^{-2} \right)^{-1} \right. \quad (22)$$
$$\left. \cdot I_3 \right),$$

where $n_{k,c}$ are the number of foci in some cluster k and subcluster c and $\overline{(s_{rj} - p_r)}$ denotes the mean of the observed data after adjusting for group effect. The value for p_r for those groups (studies) in group cluster c (different c notation than subclusters above) may be sampled from

$$p_c \mid \cdot \sim TMVN_3 \left(\left(\sigma_p^{-2} + n_c\sigma^{-2} \right)^{-1} \right.$$
$$\left. \cdot \left(\sigma_p^{-2}\boldsymbol{\mu_p} + n_c\sigma^{-2}\overline{(s_{rj} - \theta_{rj})} \right), \left(\sigma_p^{-2} + n_c\sigma^{-2} \right)^{-1} I_3, \quad (23)$$
$$\left. -l, l \right),$$

where n_c is the number of groups in cluster c and similarly $\overline{(s_{rj} - \theta_{rj})}$ denotes the mean of the observed data after adjusting for individual effect.

2.4. Determining the Clusters. To estimate the number of clusters and the center of each cluster and cluster assignment, we implement the same least-squared Euclidean distance method introduced in [15] and used previously in our work in [10] and reiterated below. This method draws inferences on clusters based on a set of converged MCMC iterations and chooses one iteration as the final estimates for the clusters and related parameters. This final MCMC iteration is selected due to its smallest Euclidean distance to the expected cluster assignments estimated based on a set of independent converged MCMC iterations. This approach incorporates all clustering information in the MCMC sampling process [15]:

(1) After the prespecified number of MCMC burn-ins, let the MCMC simulations continue for an additional W iterations. An averaged clustering matrix is then created, denoted as A, and is an $n \times n$ matrix with each block or $(i, j)^{th}$ entry denoting how often foci i and j $(i, j = 1, \ldots, n)$ are in the same cluster. Specifically, each $(i, j)^{th}$ entry is the proportion of the W iterations that two foci are in the same cluster.

(2) Let the MCMC run additional F iterations, where, for each iteration,

(a) create an $n \times n$ matrix using indicators to denote which foci are clustered together; i.e., let the

$(i, j)^{th}$ entry denote a 1 if foci i and j are in one cluster and 0 otherwise.

(b) use Euclidean distance to determine the similarity between this indicator matrix and the averaged clustering matrix A.

(3) Among the F iterations, select the iteration and respective clustering pattern, number of clusters, and parameters that produce the smallest Euclidean distance.

3. Inference

3.1. Simulation Settings. Simulations were utilized to illustrate and assess the proposed method. We assumed an fMRI metadata study setting including 50 studies, with each study containing 10 foci. Collectively these foci were simulated from three clusters centered as (x,y,z) talairach coordinates at $(1, 1, 1)^T$, $(2, 2, 2)^T$, and $(4, 4, 4)^T$ containing 150, 150, and 200 foci, respectively. It was also assumed that half the data, 250 foci or half from each individual cluster, came from one study cluster centered at $(0.1, 0.1, 0.1)^T$ and the remaining data from a second study cluster centered at $(0.4, 0.4, 0.4)^T$. These study clusters were linear shifts to cluster centers. For example, 75 foci in cluster one were centered at $(1.1, 1.1, 1.1)^T$ and the other half were centered at $(1.4, 1.4, 1.4)^T$; the study effect was a linear shift to all three dimensions from the cluster center. We made various alterations to this general setting:

(1) Normal setting: We simulated data from a multivariate normal for each cluster with respective means described above and a variance of $\Sigma = 0.002I_3$. This creates spheres with little variation and we expect the method to have the ability to correctly identify the clusters.

(2) Chi-squared (skewed) setting: The method's ability to cluster in the presence of abnormal patterns is an important factor in spatial clustering. For this setting, we applied the same scenario as in the normal setting in respect of clusters 1 and 2 but simulated cluster 3 using a chi-squared distribution with 4 degrees of freedom.

(3) Large variance setting: The last scenario is designed to assess the robustness of the method with respect to the distance between and among clusters. To this end, we applied the normal setting but considered increasing levels (referred to large1, large2, large3, and large4 settings, resp.) of Σ: $\Sigma = 0.01I_3, 0.05I_3, 0.1I_3$, and $0.2I_3$, representing gradually closer distances among clusters.

For each setting, we implemented a grid search for a single dataset to estimate values of α_p and α based on the minimization of DIC. We let precision parameter values be 0.01, 0.05, 0.1, 0.5, 1, 2, and 5. Based on α_p and α estimates, 100 MC datasets were generated with 2500 burn-in iterations, 500 working iterations to calculate the probability matrix for determining the clusters, and 100 additional iterations

TABLE 1: Simulation assessments for three foci-level clusters.

Scenario (α_p, α)	Median Num. of Clusters (SD)*	Cluster index	Average Sensitivity (SD)*	Average Specificity (SD)*	Average Accuracy (%) (SD)*
Normal (0.1, 1.0)	**IC: 3 (0.88)	1	0.96 (0.13)	1 (0)	0.96 (0.13)
		2	0.96 (0.13)	1 (0)	
		3	0.96 (0.13)	1 (0)	
	**SC: 2 (0.32)	1	0.93 (0.26)	1 (0.04)	0.96 (0.13)
		2	1 (0.04)	0.93 (0.26)	
Chi-squared (0.05, 0.1)	IC: 19 (2.17)	1	0.94 (0.09)	0.91 (0.15)	0.54 (0.13)
		2	0.74 (0.43)	0.96 (0.02)	
		3	0.09 (0.03)	0.94 (0.16)	
	SC: 2 (0.56)	1	0.76 (0.37)	0.9 (0.1)	0.82 (0.18)
		2	0.88 (0.11)	0.78 (0.36)	
Large 1 (0.05, 0.1)	IC: 3 (0.82)	1	0.96 (0.13)	1 (0)	0.96 (0.13)
		2	0.96 (0.13)	1 (0)	
		3	0.96 (0.13)	1 (0)	
	SC: 2 (0.29)	1	0.92 (0.27)	1 (0)	0.96 (0.14)
		2	1 (0)	0.92 (0.27)	
Large 2 (0.01, 0.5)	IC: 3 (0.77)	1	0.99 (0.04)	1 (0)	0.99 (0.02)
		2	0.99 (0.03)	1 (0)	
		3	1 (0.01)	1 (0)	
	SC: 2 (0.14)	1	1 (0.02)	1 (0.01)	1 (0.01)
		2	1 (0.01)	1 (0)	
Large 3 (0.05, 0.05)	IC: 5 (1.25)	1	0.91 (0.13)	1 (0)	0.92 (0.07)
		2	0.88 (0.16)	1 (0)	
		3	0.96 (0.09)	1 (0)	
	SC: 2 (0.37)	1	0.97 (0.06)	0.98 (0.03)	0.97 (0.04)
		2	0.98 (0.03)	0.98 (0.03)	
Large 4 (0.01, 2.0)	IC: 9 (1.3)	1	0.66 (0.14)	0.99 (0.01)	0.62 (0.09)
		2	0.33 (0.08)	0.99 (0.01)	
		3	0.8 (0.17)	1 (0)	
	SC: 1 (0.36)	1	0.04 (0.17)	0.99 (0.05)	0.51 (0.08)
		2	0.98 (0.07)	0.05 (0.18)	

*SD: standard deviation across 100 MC replicates; **IC: individual foci cluster; SC: study effect clusters.

to infer the number of clusters and individual foci cluster centers.

Model assessment consists of three evaluations: sensitivity, specificity, and accuracy. Sensitivity and specificity are defined by their generic definitions, the proportion of foci that are correctly assigned to their simulated cluster, and the proportion of foci that are correctly not assigned their nonsimulated cluster. Accuracy is defined as the percentage of foci that are correctly clustered. Note that the definition of accuracy takes into account both true positive and true negatives. In addition to our methodology, we applied a very common existing clustering approach for continuous data, K-means, to our simulation settings. Although this method cannot adjust for additional covariates, it allows for a comparison to existing methods. Lastly, to highlight the advantage of using a mixture of DPs over existing clustering

approaches, we applied our approach, a revised version of our approach using a mixture of multivariate normal distributions rather than DPs, and Kmeans to the normal and chi-squared simulation scenarios. As the emphasis was on clustering performance, the group effect was assumed to be known for these two settings.

3.2. Simulation Results. Table 1 summarizes the findings on the three foci-level cluster identifications and the quality of the identified clusters. The proposed method gives high sensitivities and specificities across all scenarios. The accuracy of cluster assignment overall is higher than 90% when the variation in the data is relatively small and only dropping once clusters were large enough to overlap (scenario Large 4). The proposed methodology was also accurate at identifying the correct number of clusters as indicated by the median

TABLE 2: Kmeans simulation assessments for three foci-level clusters.

Scenario:	Median Num. of Clusters (SD)∗	Cluster index	Average Sensitivity (SD)∗	Average Specificity (SD)∗	Average Accuracy (%) (SD)∗
Normal	5 (0.95)	1	0.73 (0.29)	0.98 (0.09)	
		2	0.72 (0.3)	0.98 (0.08)	0.7 (0.12)
		3	0.67 (0.26)	1 (0)	
Chi-squared	10 (0.5)	1	0.58 (0.19)	0.99 (0.01)	
		2	0.86 (0.28)	0.92 (0.03)	0.5 (0.09)
		3	0.16 (0.08)	0.98 (0.11)	
Large 1	5 (0.98)	1	0.68 (0.33)	1 (0.04)	
		2	0.76 (0.26)	0.96 (0.12)	0.7 (0.13)
		3	0.67 (0.26)	1 (0)	
Large 2	7 (1.29)	1	0.61 (0.27)	0.99 (0.06)	
		2	0.61 (0.27)	0.99 (0.06)	0.58 (0.11)
		3	0.54 (0.27)	1 (0)	
Large 3	9 (0.94)	1	0.49 (0.19)	1 (0)	
		2	0.47 (0.17)	1 (0)	0.46 (0.08)
		3	0.43 (0.24)	1 (0)	
Large 4	10 (0.57)	1	0.43 (0.16)	1 (0.01)	
		2	0.42 (0.14)	1 (0.01)	0.41 (0.07)
		3	0.38 (0.2)	1 (0)	

∗SD: standard deviation across 100 MC replicates.

number of individual and study clusters. In comparison, results from the Kmeans approach indicate relatively lower statistics in sensitivity and specificity (Table 2). The accuracy is around 70%, when the variations in the data are relatively small. However, compared to the proposed method, the Kmeans approach often inferred a higher number of clusters as indicated by the larger median number of clusters. The computation time of a single dataset for the mixture of DPs took, on average, 7-8 hours on a high performance computer (Dell cluster with 88 compute nodes, 3120 total central processing unit cores, 20664 Giga-bytes of RAM, and 61440 total graphic processing unit cores).

After removing study effects, the comparison between the mixture of DPs, mixture of normal, and Kmeans is as expected (Table 3). Both mixtures performed exceptionally well at identifying the three normally distributed clusters while the Kmeans performance was adequate with an overall accuracy of 80% (compared to 100% for both mixtures). Once the data deviated from normality, the mixture of normals approach was unable to differentiate the clusters resulting in low accuracy (32%). The Kmeans approach performed similarly to the mixture of DPs with both approaches having low sensitivities for the third cluster which was skewed but with the mixture of DPs resulting in superior accuracy. The mixture of DPs was able to differentiate clusters 1 and 2 as indicated by 99% sensitivity and 100% specificity measures but tended to "overcluster" cluster 3 into smaller clusters as indicated by the large median number of clusters, 14% sensitivity, and 100% specificity. In regard to overall accuracy,

the mixture of DPs outperformed the mixture of normals and Kmeans when the data are skewed.

4. Real Data Application

For this application, we applied the proposed method to a meta-analysis dataset. Constructed originally in [16], this data consists of a total of 162 neuroimaging publications, of which 57 were PET and 105 were fMRI. Among these 162 publications, there were 437 contrasts or studies. Only those foci that were deemed significantly activated by their study specific criteria were included for a total of 2,478 foci. Summary statistics for this data can be seen in Tables 4 and 5.

As with the simulation studies, grid search and DIC were used to estimate values for α_p and α. Potential precision parameters values were 0.01, 0.05, 0.1, 0.5, 1, 2, 5, and 7.5. Each combination was performed over 2,600 iterations, 2,000 of those for burn-in, 500 for the probability matrix calculation, and final 100 to infer individual clusters and their centers. To assist with the magnitude of the likelihood calculations, the data was scaled down by 10.

It was found that the precision parameter combination of $\alpha_p = 0.05$ and $\alpha = 1$ produced the smallest DIC. Convergence, with the initial 2,000 discarded, was checked based on trace plots. Based on the proposed method, we identified four study clusters and 14 individual foci clusters. A single DIC setting for this data took, on average, 72 hours to run on the HPC. The break down of the 14

TABLE 3: Comparison of approaches for selected simulated settings assuming study effect is known.

Setting	Approach	Median Num. of Clusters (SD)*	Cluster index	Average Sensitivity (SD)*	Average Specificity (SD)*	Average Accuracy (%) (SD)*
Normal	DP	3 (0)	1	1 (0)	1 (0)	1 (0)
			2	1 (0)	1 (0)	
			3	1 (0)	1 (0)	
	Mixture	3 (0)	1	1 (0)	1 (0)	1 (0)
			2	1 (0)	1 (0)	
			3	1 (0)	1 (0)	
	Kmeans	3 (0.6)	1	0.81 (0.36)	0.93 (0.16)	0.8 (0.22)
			2	0.79 (0.38)	0.94 (0.15)	
			3	0.79 (0.25)	1 (0)	
Chi-squared	DP	16 (1.76)	1	0.99 (0.01)	0.99 (0.01)	0.65 (0.01)
			2	0.99 (0.01)	0.97 (0.01)	
			3	0.14 (0.03)	1 (0)	
	Mixture	1 (0.29)	1	0.06 (0.24)	0.98 (0.11)	0.32 (0.1)
			2	0.95 (0.22)	0.06 (0.24)	
			3	0.05 (0.21)	1 (0)	
	Kmeans	10 (0.22)	1	0.99 (0.06)	0.98 (0.01)	0.62 (0.1)
			2	0.87 (0.34)	0.93 (0.03)	
			3	0.16 (0.06)	0.94 (0.17)	

*SD: standard deviation across 100 MC replicates.

TABLE 4: Descriptive statistics*.

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
Number of foci per pub.	1.00	5.75	10.00	15.11	17.25	110.00
Number of foci per study	1.00	2.00	4.00	5.67	7.00	47.00
Number of subjects per pub.	4.00	9.00	11.00	12.26	14.00	40.00
Number of studies per pub.	1.000	1.000	2.000	2.67	4.000	12.000

*Min: minimum, 1st Qu: 25% percentile, 3rd Qu: 75% percentile, Max: maximum, pub: publication.

TABLE 5: Frequency of emotions.

Emotions	Frequency of studies (% of total studies)	Frequency of foci (% of total foci)
aff*	175 (40.05%)	881 (35.55%)
anger	26 (5.95 %)	166 (6.7%)
disgust	44 (10.07%)	337 (13.6%)
fear	68 (15.56%)	367 (14.81%)
happiness	36 (8.24%)	178 (7.18%)
mixed	41 (9.38%)	195 (7.87%)
sadness	45 (10.3%)	348 (14.04%)
surprise	2 (0.46%)	6 (0.24%)
Total	437	2478

*aff: affective.

individual foci clusters by center location, brain location, foci frequency, and study frequency can be seen in Table 6. The frequency of each foci-associated emotion within each of the 14 clusters can be seen in Table 7. The affective emotion was dominating in all clusters with fear being the second dominating emotion in clusters 1, 2, 3, 11, and 13, disgust in clusters 5 and 14, sadness in clusters 6 and 10, and a mixture of emotions in the remaining clusters. When only focusing on those foci that fell within known brain regions of interest, as seen in Table 8, the dominating emotion, other

Table 6: Meta-data cluster results.

Cluster Centers	Brian Regions	Cluster Index	# of foci per cluster (% of total foci)	# of studies per cluster (% of all studies)
(-9.11, -14.45, 1.4)	Temporal Mid L	1	1615 (65.17)	386 (88.33)
(-0.83, -4.64, 0.54)	Temporal Mid L	2	188 (7.59)	139 (31.81)
(16.75, -10.63, -11.33)	Temporal Mid R	3	148 (5.97)	117 (26.77)
(3.48, -5.52, -2.15)	Temporal Inf L	4	102 (4.12)	84 (19.22)
(2.08, -4.61, -0.14)	Temporal Mid R	5	77 (3.11)	67 (15.33)
(0.87, -7.7, -0.63)	Temporal Pole Sup L	6	67 (2.7)	59 (13.5)
(0.99, -6.33, 2.62)	Cerebelum 6 R	7	55 (2.22)	43 (9.84)
(2.3, -5.89, 0.32)	NA	8	54 (2.18)	46 (10.53)
(0.43, -6.59, 0.6)	Postcentral L	9	41 (1.65)	36 (8.24)
(-0.18, -4.99, 0.11)	Cerebelum 6 R	10	39 (1.57)	38 (8.7)
(1.69, -5.25, 1.43)	Temporal Sup R	11	38 (1.53)	34 (7.78)
(-0.64, -5.93, 1.77)	Postcentral L	12	22 (0.89)	20 (4.58)
(1.29, -5.28, 1.23)	Precentral R	13	18 (0.73)	17 (3.89)
(0.45, -5.31, -0.87)	Occipital Inf R	14	14 (0.56)	14 (3.2)

R: right hemisphere, L: left hemisphere.

than affective, was sadness, fear, and disgust, respectively. When compared to the number of clusters identified by the spatial Cox Point process (53) and Kmeans (20) performed in [10], fewer clusters were identified with our current application. It should be noted that this particular data does not visually indicate distinct clusters and is closer to a more uniform distribution throughout the brain which may lead to an inaccurate number of identifiable clusters. However, given the results from our previous analysis in [10] and findings in simulation studies, it is possible that the clusters formed were rather subtle and actually might not be distinct enough.

5. Conclusion and Discussion

Modeling the realization of observed foci as a linear association of study effect and individual foci cluster effect with a multivariate normal random error was motivated by the limitation of the spatial Cox process to statistically distinguish between a cluster and a mode or peak of a cluster. The overall aim remained to identify activated regions within the brain using fMRI coordinate-based metadata. By modeling the data in this fashion, it was hopeful that the distribution could statistically differentiate between clusters and modes of clusters while retaining the flexibility and robustness to mimic the behavior of the data.

Simulation studies demonstrated that the method can fit data generated from normal or abnormal distributions. Furthermore, it was able to identify clusters within covariates while retaining the integrity to identify individual clusters. Both the proposed method and Kmeans were unable to correctly identify clusters when they were large and overlapped and both the mixture of normals and mixture of DPs performed poorly at identifying a cluster severely skewed.

When applied to a fMRI metadataset, the method identified a relatively low number of clusters. Given the low sensitivity findings in the simulated study with high noise, it can be concluded that this data had a high likelihood of being too broad. When the same data was analyzed with the spatial Cox process, the difference in the results was extreme. Not only were the number of clusters substantially less, but also none of the cluster centers identified from the proposed method came close to those identified in the first method. It is worth mentioning that the meta-analysis data is not distinctly grouped and is more uniformly distributed throughout the brain and perhaps the model used did not provide the best fit.

The primary advantage to this method, besides its flexibility, is its ability to describe irregular spatial patterns and its sampling design to statistically differentiate clusters. Because of its adaptable nature, this model can also adjust for any covariate(s) of interest. However, based on simulation studies and the fMRI metadata application, the proposed method tends to be too insensitive and has a difficult time identifying clusters when data are not distinctly differentiated. A potential limitation in the approach is that each DP within the mixture was assumed to have the same precision parameter. It was noted that, during the simulation studies when the mixture of DPs was attempting to fit the Chi-squared simulations (without study effect), it was overclustering the skewed cluster. However, when the precision parameter was smaller, the identifiability of cluster 3 became more accurate but became more inaccurate for clusters 1 and 2. Thus, to further improve flexibility and accuracy for this approach when data is skewed, each DP potentially requires its own unique precision parameter. Furthermore, this method's clustering ability is limited by the identification of study effects which may be improved by implementing stronger restrictions or could be an effect of having multiple DPs. Our future work will focus on these issues, allowing study effect to be random rather than a fixed effect, and identifying if a large number of DPs within the model is indeed a limitation.

TABLE 7: Breakdown of emotions and their frequencies by individual foci cluster∗.

	Cluster Index: Total foci in that cluster				
	Emotion Frequency of emotion (% of total cluster foci)				
Cluster: 1	1615	**Cluster: 2**	188	**Cluster: 3**	148
aff	562 (34.8)	aff	57 (30.32)	aff	53 (35.81)
anger	110 (6.81)	anger	10 (5.32)	anger	12 (8.11)
disgust	225 (13.93)	disgust	18 (9.57)	disgust	22 (14.86)
fear	234 (14.49)	fear	35 (18.62)	fear	31 (20.95)
happiness	121 (7.49)	happiness	17 (9.04)	happiness	8 (5.41)
mixed	131 (8.11)	Mixed	18 (9.57)	mixed	8 (5.41)
sadness	228 (14.12)	sadness	33 (17.55)	sadness	13 (8.78)
surprise	4 (0.25)			surprise	1 (0.68)
Cluster: 4	102	**Cluster: 5**	77	**Cluster: 6**	67
aff	44 (43.14)	aff	35 (45.45)	aff	22 (32.84)
anger	6 (5.88)	anger	3 (3.9)	anger	4 (5.97)
disgust	15 (14.71)	disgust	13 (16.88)	disgust	5 (7.46)
fear	9 (8.82)	fear	5 (6.49)	fear	14 (20.9)
happiness	9 (8.82)	happiness	5 (6.49)	happiness	6 (8.96)
mixed	4 (3.92)	Mixed	7 (9.09)	Mixed	4 (5.97)
sadness	15 (14.71)	sadness	9 (11.69)	sadness	11 (16.42)
				surprise	1 (1.49)
Cluster: 7	55	**Cluster: 8**	54	**Cluster: 9**	41
aff	21 (38.18)	aff	19 (35.19)	aff	13 (31.71)
anger	7 (12.73)	anger	2 (3.7)	anger	5 (12.2)
disgust	7 (12.73)	disgust	9 (16.67)	disgust	8 (19.51)
fear	6 (10.91)	fear	7 (12.96)	fear	8 (19.51)
happiness	3 (5.45)	mixed	9 (16.67)	happiness	2 (4.88)
mixed	4 (7.27)	sad	8 (14.81)	mixed	1 (2.44)
sadness	7 (12.73)			sad	4 (9.76)
Cluster: 10	39	**Cluster: 11**	38	**Cluster: 12**	22
aff	15 (38.46)	aff	20 (52.63)	aff	9 (40.91)
anger	2 (5.13)	anger	1 (2.63)	anger	2 (9.09)
disgust	5 (12.82)	disgust	2 (5.26)	disgust	3 (13.64)
fear	3 (7.69)	fear	7 (18.42)	fear	3 (13.64)
happiness	3 (7.69)	happiness	2 (5.26)	mixed	2 (9.09)
mixed	4 (10.26)	sad	6 (15.79)	sad	3 (13.64)
sadness	7 (17.95)				
Cluster: 13	18	**Cluster: 14**	14		
aff	5 (27.78)	aff	6 (42.86)		
anger	2 (11.11)	disgust	3 (21.43)		
disgust	2 (11.11)	happiness	1 (7.14)		
fear	5 (27.78)	mixed	2 (14.29)		
happiness	1 (5.56)	sad	2 (14.29)		
mixed	1 (5.56)				
sadness	2 (11.11)				

TABLE 8: Breakdown of emotions and their frequencies by individual foci cluster for ROI*.

Cluster Index: Total foci in that cluster					
Emotion Frequency of emotion (% of total cluster foci)					
Cluster: 1	489	**Cluster: 2**	63	**Cluster: 3**	54
aff	217 (44.38)	aff	29 (46.03)	aff	21 (38.89)
anger	20 (4.09)	anger	3 (4.76)	anger	3 (5.56)
disgust	68 (13.91)	disgust	8 (12.7)	disgust	4 (7.41)
fear	63 (12.88)	fear	13 (20.63)	fear	8 (14.81)
happiness	28 (5.73)	happiness	2 (3.17)	happiness	6 (11.11)
mixed	22 (4.5)	mixed	4 (6.35)	mixed	5 (9.26)
sadness	71 (14.52)	sadness	4 (6.35)	sadness	7 (12.96)
Cluster: 4	34	**Cluster: 5**	25	**Cluster: 6**	20
aff	15 (44.12)	aff	13 (52)	aff	8 (40)
anger	1 (2.94)	anger	1 (4)	disgust	1 (5)
disgust	4 (11.76)	disgust	4 (16)	fear	4 (20)
fear	3 (8.82)	fear	1 (4)	happiness	1 (5)
happiness	1 (2.94)	happiness	2 (8)	mixed	1 (5)
mixed	1 (2.94)	mixed	2 (8)	sadness	5 (25)
sadness	9 (26.47)	sadness	2 (8)		
Cluster: 7	19	**Cluster: 8**	17	**Cluster: 9**	16
aff	6 (31.58)	aff	6 (35.29)	aff	9 (56.25)
anger	1 (5.26)	anger	2 (11.76)	fear	3 (18.75)
disgust	4 (21.05)	disgust	3 (17.65)	happiness	2 (12.5)
mixed	2 (10.53)	happiness	1 (5.88)	sadness	2 (12.5)
sadness	6 (31.58)	mixed	1 (5.88)		
		sadness	4 (23.53)		
Cluster: 10	15	**Cluster: 11**	9	**Cluster: 12**	7
aff	6 (40)	aff	3 (33.33)	aff	3 (42.86)
anger	1 (6.67)	anger	1 (11.11)	anger	1 (14.29)
disgust	2 (13.33)	disgust	2 (22.22)	disgust	1 (14.29)
fear	2 (13.33)	fear	1 (11.11)	fear	2 (28.57)
happiness	1 (6.67)	sadness	2 (22.22)		
mixed	2 (13.33)				
sadness	1 (6.67)				
Cluster: 13	6	**Cluster: 14**	5		
aff	3 (50)	aff	3 (60)		
disgust	1 (16.67)	happiness	1 (20)		
sadness	2 (33.33)	sadness	1 (20)		

*ROI: region of interest; aff: affective.

Data Availability

Data is not currently publicly available but available upon request from Professor Tor D. Wager at the University of Colorado.

Conflicts of Interest

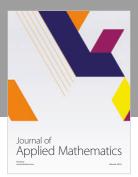No conflicts of interest exist for any author.

Acknowledgments

References

[1] M. Stephens, "Bayesian analysis of mixture models with an unknown number of components—an alternative to reversible jump methods," *The Annals of Statistics*, vol. 28, no. 1, pp. 40–74, 2000.

[2] M. Aitkin and R. Healey, "Estimation and hypothesis testing in finite mixture models. Journal of the Royal Statistical Society," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 67–75, 1985.

[3] G. J. McLachlan and K. E. Basford, *Mixture Models: Applications to Clustering*, Marcel Dekker, New York, NY, USA, 1988.
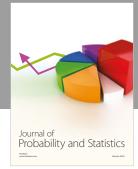
[4] M. A. T. Figueiredo and A. K. Jain, "Unsupervised learning of finite mixture models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 3, pp. 381–396, 2002.

[5] J.-M. Marin, K. Mengersen, and C. . Robert, "Bayesian modelling and inference on mixtures of distributions," in *Bayesian thinking: modeling and computation*, vol. 25 of *Handbook of Statist.*, pp. 459–507, Elsevier/North-Holland, Amsterdam, 2005.

[6] P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika*, vol. 82, no. 4, pp. 711–732, 1995.

[7] P. J. Green and D. I. Hastie, "Reversible jump mcmc," *Genetics*, vol. 1550, no. 3, pp. 1391–1403, 2009.

[8] M. D. Escobar and M. West, "Bayesian density estimation and inference using mixtures," *Journal of the American Statistical Association*, vol. 90, no. 430, pp. 577–588, 1995.

[9] C. E. Antoniak, "Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems," *The Annals of Statistics*, vol. 2, pp. 1152–1174, 1974.

[10] M. Ray, J. Kang, and H. Zhang, "Identifying Activation Centers with Spatial Cox Point Processes Using fMRI Data," *IEEE Transactions on Computational Biology and Bioinformatics*, vol. 13, no. 6, pp. 1130–1141, 2016.

[11] P. Congdon, *Bayesian statistical modelling*, vol. 704, John Wiley & Sons, 2007.

[12] D. J. Spiegelhalter, N. G. Best, B. P. Carlin, and A. van der Linde, "Bayesian measures of model complexity and fit," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 64, no. 4, pp. 583–639, 2002.

[13] C. Preston, "Spatial birth-and-death processes," *Advances in applied probability*, vol. 70, no. 03, pp. 405–408, 1975.

[14] R. M. Neal, "Markov chain sampling methods for Dirichlet process mixture models," *Journal of Computational and Graphical Statistics*, vol. 9, no. 2, pp. 249–265, 2000.

[15] D. B. Dahl, "Model-based clustering for expression data via a dirichlet process mixture model," *Bayesian Inference for Gene Expression and Proteomics*, pp. 201–218, 2006.

[16] H. Kober, L. F. Barrett, J. Joseph, E. Bliss-Moreau, K. Lindquist, and T. D. Wager, "Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies," *NeuroImage*, 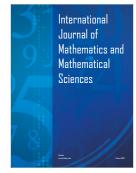vol. 42, no. 2, pp. 998–1031, 2008.