



HHS Public Access

Author manuscript

IEEE/ACM Trans Comput Biol Bioinform. Author manuscript; available in PMC 2017 January 24.

Published in final edited form as:

IEEE/ACM Trans Comput Biol Bioinform. 2016 ; 13(6): 1130–1141. doi:10.1109/TCBB.2015.2510007.

Identifying Activation Centers with Spatial Cox Point Processes Using FMRI Data

Meredith Ray,

University of Memphis, School of Public Health, Division of Epidemiology, Biostatistics, and Environmental Health, Memphis, TN, 38111

Jian Kang, and

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48105

Hongmei Zhang [Member IEEE]

University of Memphis, School of Public Health, Division of Epidemiology, Biostatistics, and Environmental Health, Memphis, TN, 38111

Abstract

We developed a Bayesian clustering method to identify significant regions of brain activation. Coordinate-based meta data originating from functional magnetic resonance imaging (fMRI) were of primary interest. Individual fMRI has the ability to measure the intensity of blood flow and oxygen to a location within the brain that was activated by a given thought or emotion. The proposed method performed clustering on two levels, latent foci center and study activation center, with a spatial Cox point process utilizing the Dirichlet process to describe the distribution of foci. Intensity was modeled as a function of distance between the focus and the center of the cluster of foci using a Gaussian kernel. Simulation studies were conducted to evaluate the sensitivity and robustness of the method with respect to cluster identification and underlying data distributions. We applied the method to a meta data set to identify emotion foci centers.

Index Terms

Bayesian; clustering; fMRI; meta analysis

1 Introduction

Functional neuroimaging is a new branch within the neurosciences that is being credited with advancing our understanding of the human brain and its processes in neuroscientific research related to public health. Accurately identifying activated foci or regions in the brain in association with an outcome of interest is crucial in terms of disease prediction and prevention. Functional magnetic resonance imaging (fMRI) is the most widely used method for this type of study. FMRI is able to measure changes in tissue perfusion, blood oxygenation, and blood volume [1]. The amount and location of these changes are measured when subjects are under some situational environment or experiment that provokes a thought, emotion, or action. Brain activation recorded in fMRI happens before, during, and after the expected stimulation and thus gives rise to background noise or simulation occurring that is not due to the stimulant and therefore must decipher which areas of

activation are statistically greater than the general mass activity or noise [2]. Furthermore, because of the high cost, fMRI studies tend to have very small number of participants (less than 20) which causes inflated type II error (low power) and lack of reproducibility [3]. To increase sample size and testing power, meta analyses are being implemented to summarize these smaller individual studies ([4], [5], [6], [7], [8]). Two types of meta analyses are commonly used, intensity based meta analysis (IBMA) and coordinate based meta analysis (CBMA) [7], corresponding to the two types of individual fMRI data with one represented a statistical image/map (intensity based) and the other as a coordinate system (coordinate based), respectively. Our primary focus is on the more commonly used coordinate based meta analysis data [1].

Meta data of CBMA only supplies the (x, y, z) coordinates or foci (focus-single) of peak activations. The activation magnitude may or may not be given [7]. The overall aim of CBMA is to identify areas within the brain that are statistically activated during a given stimulant. A variety of statistical methods for analyzing CBMA have been proposed: activation likelihood estimation (ALE) ([6], [9], [10], [11], [12]), kernel density analysis (KDA) ([5], [13], [14]), and a hierarchical spatial point process model proposed in [8].

Activation likelihood estimation (ALE), based on the assumption that all foci are independent, divides the brain into a smaller cubes called voxels. The probability of a foci being within a particular voxel is modeled using a 3-D Gaussian function. All probabilities are summed for each voxel resulting in an activation likelihood estimate, which is interpreted that at least one of the activated foci is within that voxel ([6], [9], [10]). Advancements have been made to improve this original work such as methods to determine the probability distribution size, estimating clusters between studies instead of within, and controlling for false discovery ([10], [11], [12]). The need for these advancements indicates the results can vary due to voxel size, Gaussian kernel standard deviation, and controlling for multiple testing.

Kernel density analysis (KDA), similar to ALE in that it treats each foci independent and divides the brain into voxels, implements a smoothing kernel which is a spherical indicator function with a given radius for each foci. After the smoothing, a smoothed histogram reflects the estimated density of each activation location in each voxel ([13], [15]). Advancements have been made to treat studies as a unit of analysis rather than foci (multilevel KDA). In this case, the proportion of studies becomes the focus rather than the number of foci that are activated in a specific region of the brain [14]. Activation peaks are nested within studies when the test statistics are calculated which allows each study to contribute proportionally ([5], [14]). Although this multilevel KDA made significant improvements, the analysis is still driven by voxel size, radius size for the smoothing spherical kernel, and an underlying distribution assumption [7].

The hierarchical spatial point process model proposed by Kang et al. [8] is formulated from a marked Cox process and identifies population centers that represent the most likely activation location of a foci across studies. The model is comprised of three levels. The first layer models the individual foci, assumed independent, by clustering with an independent cluster process controlled by a random intensity function. At this level, two types of foci are

considered: singly and multiply reported, which were assigned a 'mark' as to identify the specific types. The second level models the latent study activation center, again assumed independent and clustered by an independent cluster process and driven by a random intensity function. These two layers are both assumed conditional on the realization of latent study activation centers or latent population centers, respectively, and normally distributed. Also, both layers allow for individual foci and study centers to be singleton clusters and model this by an independent homogeneous Cox process. The third and final level models the latent population centers with a homogeneous Cox process controlled by a homogeneous random intensity. This method has the power to identify clusters at different levels with final focus on the population level. However, assuming the realization of both the foci and studies to be normally distributed limits the model to only identify clusters of a certain distribution or shape and therefore can miss or mis-cluster foci.

In this article, we take a different route and propose a Bayesian hierarchical spatial Cox point process model [16] to identify individual foci clusters while adjusting study effects using meta data. The method incorporates two nesting Dirichlet processes (DPs) [17]. To adjust for study effect, we implemented a Dirichlet process because of its ability to describe abnormal studies effects. The Gaussian kernel is utilized to measure the contribution of a focus with respect to a cluster center to the intensity of the Cox point process. The foci clusters are identified after adjusting for study effect and through a Dirichlet process as well. Via these two DPs, we gain the potential to describe possibly more complicated patterns of spatial effects on the intensity. This approach is different from the method by [8] in that the focus there was at the population level, while the proposed method is on clustering individual foci while adjusting study effect. Both ALE and KDA methods, on the other hand, are voxel based approaches with no focus on clustering and are conditional on those specific brain locations. Our method is a method for clustering. Compared to the existing clustering methods (such as the K-means approach), the proposed approach is able to incorporate prior knowledge of study effect into the clustering process, has the potential to incorporate other factors, and can identify foci centers with irregular shapes.

Model details including prior, joint posterior, and conditional posterior distributions and the algorithm applied to exercise the Dirichlet process, are discussed in section 2. Section 3 presents 7 simulation studies and section 4 describes an application of the method to a meta data set, which was discussed in [8] and [5]. Finally, section 5 presents a discussion of the findings, thoughts on our model, and future research.

2 Model

2.1 Notation

The general spatial Poisson point process is defined on a space M , such that $M \subseteq \mathbb{R}^3$ and driven by a locally integrable intensity function, $\nu: M \rightarrow [0, \infty)$ and $\int_R \nu(\xi) d\xi < \infty$ for all bounded $R \subseteq M$ [16]. Define an intensity measure $\mu(R) = \int_R \nu(\xi) d\xi$, which is also locally finite for bounded $R \subseteq M$ and diffuse, $\mu(\xi) = 0$ for $\xi \in M$ [16]. By letting this intensity function be a realization of a nonnegative random field, $Z = \{Z(\xi) : \xi \in M\}$, this extends to a Cox process. Straightforwardly, by restricting this Cox process to $|R| < \infty$ the following density function can be attained [16]:

$$f(x) = \exp \left\{ |R| - \int_R Z(\xi) d\xi \right\} \prod_{\xi \in x} Z(\xi).$$

Due to the process' ability to model and cluster within spatial randomness, we can directly apply this model to any spatial data but in particular, CBMA data. Let the intensity measure be driven by multiplicative of a Gaussian kernel which is bounded by the brain space B , such that $B \subseteq \mathbb{R}^3$ and $|B| < \infty$. Furthermore, let $s_{ij} = (x, y, z)$ denote a single focus which represents a talairach or Montreal Neurological Institute (MNI) coordinate in the brain, for sub-study j in study i , $i = 1, \dots, I, j = 1, \dots, J_i$. We have $\sum_{i=1}^I J_i = n$, where n is the total number of observed foci. Denoted by $\mathbf{s} = s_{1,1} \dots s_{I,J_I}$ represents all foci in the CBMA study such that $\mathbf{s} \subset B$. The density function above can now be redefined as:

$$f(\mathbf{s}) = E \left[\exp \left\{ |B| - \int_B \exp(a_{ij} K(s_{ij})) ds_{ij} \right\} \prod_{s_{ij} \in \mathbf{s}} K(s_{ij}) \right] \\ \propto \exp \left\{ |B| - \int_B \exp(a_{ij} K(s_{ij})) ds_{ij} \right\} \prod_{s_{ij} \in \mathbf{s}} K(s_{ij}),$$

where a_{ij} allows multiplicative focus impact on the intensity. We let the contribution of each focus be comprised of two components: expected 3-D individual focus effect (θ_{ij}) and expected 3-D study effect (p_i). Consequently, the Gaussian kernel is defined as $K(s_{ij}) = \exp(-\|s_{ij} - p_i - \theta_{ij}\|^2 / \rho)$, where ρ is a scale parameter. To avoid the unidentifiability problem due to the inclusion of a_{ij} , we set $\rho = 1$. This then yields the probability density function:

$$f(\mathbf{s}) = \exp \left[- \int_B \exp \left(a_{ij} \exp \left\{ - \frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right) ds_{ij} \right] \times \prod_{s_{ij} \in \mathbf{s}} \exp \left[a_{ij} \exp \left\{ - \frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right] \\ \approx \exp \left[- \frac{\sum_B \exp \left(a_{ij} \exp \left\{ - \frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right)}{n} + \sum_{s_{ij} \in \mathbf{s}} a_{ij} \exp \left\{ - \frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right].$$

To identify foci clusters and centers, we assume each realization s_{ij} is from the following mixture

$$f(\mathbf{s}) \propto \sum_{c=1}^C \exp \left(- \int_B \exp \left[a_c \exp \left\{ a_c \exp \left(- \frac{\|s_{ij} - p_i - \theta_c\|^2}{\rho} \right) \right\} \right] ds_{ij} \right),$$

after adjusting for study effect. To infer the clusters and their centers, we implement a fully Bayesian approach. This approach will also allow us to incorporate prior knowledge on the strength of study effect, which enables a proper adjustment for study effect and avoids biased inferences on the locations of cluster centers.

2.2 Prior and hyperprior distributions

Following the standard Bayesian framework, all estimable parameters are assigned a prior distribution. We assume study effect, p_i follows distribution G_1 , which is generated from a Dirichlet process (DP), $p_i \sim G_1$ and $G_1 \sim DP(\alpha_1, G_{01})$ where α_1 is a precision parameter and G_{01} is a base distribution. The precision parameter controls the aggregation of G_1 around G_{01} [17]. This setting results in a conditional prior for p_i ,

$$p_i | p^{(i)} \sim \frac{\alpha_1}{(\alpha_1 + I - 1)} G_{01}(p_i) + \frac{1}{(\alpha_1 + I - 1)} \sum_{q=1, q \neq i}^I \delta_{p_q}(p_i),$$

where $p^{(i)}$ represents all study effects except for study i and $\delta_{p_q}(p_i)$ denotes a unit point mass at $p_i = p_q$ [17]. The base distribution, G_{01} , is chosen as a 3-dimensional truncated multivariate normal distribution, $TMVN_3(\boldsymbol{\mu}, \Sigma_1, \mathbf{a}, \mathbf{b})$, where the hyper-prior distribution for $\boldsymbol{\mu}$ is assumed to be multivariate normal, $\boldsymbol{\mu} \sim MVN_3(\mathbf{m}_0, \sigma_\mu^2 I_3)$ with $\mathbf{m}_0 = 0$ and I_3 being a 3-dimensional identity matrix.

Parameter σ_μ^2 can take small values, e.g., $\sigma_\mu^2 = 0.1$, assuming an a priori minimal study effect. Note that \mathbf{m}_0 can choose a different value based on prior knowledge on study effect.

Parameters $\sum_1 = \sigma_1^2 I_3$ and σ_1^2 follows an inverse gamma distribution with shape and scale parameters e and f , i.e., $\sigma_1^2 \sim IG(e, f)$, $\mathbf{a} = r_a(|R_x|, |R_y|, |R_z|)'$ representing the lower limit and $\mathbf{b} = -r_b(|R_x|, |R_y|, |R_z|)'$ representing the upper limit, where r_a and r_b control the amount of truncation, set at 15%, and $|R_d|$ represents the absolute value of the range of data for coordinate d . The truncation was selected based on the assumption of small study effects and can be adjusted based on prior beliefs. The shape and scale parameters (e and f) in the inverse gamma distributions are set at both of being 0.5, $IG(0.5, 0.5)$, following the suggestion by Kass and Wasserman [18]. This prior, approximately centered at 1 and termed as “unit-information prior”, allows the conditional posterior distribution of σ_1^2 to be more influenced by data. On the other hand, when prior knowledge for study effect is clearer, an informative prior may be preferred. For θ_{ij} , we follow similar choices for p_i and assume $\theta_{ij} \sim G_2$ and $G_2 \sim DP(\alpha_2, G_{02})$, with a base distribution of $G_{02} = TMVN_3(\mathbf{c}_0, \Sigma_2, \mathbf{c}, \mathbf{d})$ where \mathbf{c}_0 takes the median of the observed data in each dimension, $\sum_2 = \sigma_2^2 I_3$, $\sigma_2^2 \sim IG(g, h)$ with $g = 0.5$ and $h = 0.5$, $\mathbf{c} = (\min(x), \min(y), \min(z))'$ and $\mathbf{d} = (\max(x), \max(y), \max(z))'$ where $\min(d)$ represents the minimum value for coordinate d and $\max(d)$ represents the maximum value to coordinate d . In both Dirichlet processes, we assume the precision parameters are known. We discuss their selection in Section 2.5. Lastly, the prior for multiplicative effect a_{ij} is conditional on identified clusters and $a_{ij} | \theta_c \sim N(0, \sigma_a^2)$ with σ_a^2 known and large.

2.3 Conditional posterior distributions and posterior computing

Posterior inference of p_i , θ_{ij} , and a_{ij} is obtained by successfully sampling values from their full conditional posterior distributions through the Markov Chain Monte Carlo (MCMC) simulations, specifically, the Gibbs sampling scheme. With s representing observed CBMA data, the joint posterior distribution of all the parameters is, up to a normalizing constant:

$$\begin{aligned}
P(p_i, \theta_{ij}, a_{ij}, \boldsymbol{\mu}, \Sigma_1, \alpha_1, \Sigma_2, \alpha_2 | \mathbf{s}) &\propto P(\mathbf{s} | p_i, \theta_{ij}, a_{ij}, \boldsymbol{\mu}, \Sigma_1, \alpha_1, \\
&\quad \times \Sigma_2, \alpha_2) \prod_i P(p_i | G_1) P(G_1 | \alpha_1, G_{01}) \\
&\quad \times P(p_i | G_{01}; \boldsymbol{\mu}, \Sigma_1) P(\boldsymbol{\mu}) P(\Sigma_1) \\
&\quad \times \prod_{c \in \mathcal{C}_i, j \in c} P(a_{ij} | \theta_{ij}) \prod_{j \in J_i} P(\theta_{ij} | G_2) \\
&\quad \times P(G_2 | \alpha_2, G_{02}) P(\theta_{ij} | G_{02}; \mathbf{c}_0, \Sigma_2) P(\Sigma_2),
\end{aligned}$$

from which full conditional posteriors for p_i , θ_{ij} , a_{ij} , along with their hyper-priors are obtained and listed below. In the following expressions, “.” represents all other parameters,

$$\begin{aligned}
&P(p_i | s_{ij}, \theta_{ij}, \cdot) \propto P(s_{ij} | p_i, \theta_{ij}, \cdot) P(\theta_{ij} | \cdot) P(p_i | G_1) P(G_1 | \alpha_1, G_{01}) \\
= \exp &\left[-\frac{\sum_B \exp \left(a_{ij} \exp \left\{ -\frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right)}{n} + \sum_{j=1}^{J_i} a_{ij} \exp \left\{ -\frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right] \\
&\quad \times \left\{ \frac{\alpha_1 TMVN_3(\boldsymbol{\mu}, \Sigma_1)}{(\alpha_1 + I - 1, \mathbf{a}, \mathbf{b})} + \frac{\sum_{q=1, q \neq i}^I \delta_{pq}(p_i)}{(\alpha_1 + I - 1)} \right\}, \\
&P(\boldsymbol{\mu} | \Sigma_1, p_i, \cdot) \propto \prod_i P(p_i | G_{01}; \boldsymbol{\mu}, \Sigma_1) P(\Sigma_1) P(\boldsymbol{\mu}) \\
= \exp &\left[-\frac{1}{2} \left\{ \prod_i (p_i - \boldsymbol{\mu})' \Sigma_1^{-1} (p_i - \boldsymbol{\mu}) + (\boldsymbol{\mu} - \mathbf{m}_0)' (\sigma_\mu^2 I_3)^{-1} (\boldsymbol{\mu} - \mathbf{m}_0) \right\} \right] \\
\boldsymbol{\mu} | \cdot &\sim TMVN_3 \left((\sigma_\mu^{-2} + I \sigma_1^{-2})^{-1} (I \sigma_1^{-2} \bar{p} + \mathbf{m}_0 \sigma_\mu^{-2}), (\sigma_\mu^{-2} + I \sigma_1^{-2})^{-1} I_3, \mathbf{a}, \mathbf{b} \right), \\
&P(\sigma_1^2 | \boldsymbol{\mu}, p_i, \cdot) \propto \prod_i P(p_i | G_{01}; \boldsymbol{\mu}, \Sigma_1) P(\boldsymbol{\mu}) P(\Sigma_1) \\
= \exp &\left[-\frac{1}{2} \left\{ \sum_i (p_i - \boldsymbol{\mu})' (\sigma_1^2 I_3)^{-1} (p_i - \boldsymbol{\mu}) \right\} + -\frac{f}{\sigma_1^2} \right], \\
\sigma_1^2 | \cdot &\sim IG \left(\frac{I}{2} + e, \frac{\sum_i \{\|p_i - \boldsymbol{\mu}\|^2\}}{2} + f \right), \\
&P(\theta_{ij} | s_{ij}, p_i, \cdot) \propto P(s_{ij} | \theta_{ij}, p_i, \cdot) P(p_i | \cdot) P(\theta_{ij} | G_2) P(G_2 | \alpha_2, G_{02}) \\
= \exp &\left[-\frac{\sum_B \exp \left(a_{ij} \exp \left\{ -\frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right)}{n} + a_{ij} \exp \left\{ -\frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} \right] \\
&\quad \times \left\{ \frac{\alpha_2 TMVN_3(\mathbf{c}_0, \Sigma_2, \mathbf{c}, \mathbf{d})}{(\alpha_2 + n - 1)} + \frac{\sum_{q=1, q \neq ij}^n \delta_{\theta q}(\theta_{ij})}{(\alpha_2 + I - 1)} \right\}, \\
&P(\sigma_2^2 | \mathbf{c}_0, \theta_{ij}, \cdot) \propto \prod_i \prod_j P(\theta_{ij} | G_{02}; \mathbf{c}_0, \sigma_2^2) P(\sigma_2^2) \\
= \exp &\left[-\frac{1}{2} \left\{ \sum_i \sum_j (\theta_{ij} - \mathbf{c}_0)' (\sigma_2^2 I_3)^{-1} (\theta_{ij} - \mathbf{c}_0) \right\} + -\frac{h}{\sigma_2^2} \right], \\
\sigma_2^2 | \cdot &\sim IG \left(\frac{n}{2} + g, \frac{\sum_{i,j} \|\theta_{ij} - \mathbf{c}_0\|^2}{2} + h \right), \\
&P(a_{ij} | s_{ij}, p_i, \theta_{ij}, c, \cdot) \propto \prod_{i,j \in c} P(s_{ij} | \theta_{ij}, p_i, \cdot) P(\theta_{ij} | c, \cdot) \times P(p_i | \cdot) P(a_{ij}) \\
= \exp &\left[-\frac{\sum_B \exp(a_{ij})}{n} \times \frac{\exp \left\{ -\frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\}}{n} + \sum_{i,j \in c} a_{ij} \exp \left\{ -\frac{\|s_{ij} - p_i - \theta_{ij}\|^2}{\rho} \right\} - \frac{(a_{ij})^2}{2\sigma_a^2} \right].
\end{aligned}$$

Non-standard conditional posteriors, a_{ij} , are updated using the Metropolis-Hastings steps in the Gibbs sampler. The parameters p_i and θ_{ij} are sampled via using an algorithm introduced in [19] that is discussed below.

The algorithm, specifically algorithm 8, noted in [19], is appropriate for models with non-conjugate priors. It introduces m auxiliary parameters to represent potential values for parameters of interest, i.e., p_i or θ_{ij} , that are not associated with any other observations [19]. The original algorithm for updating cluster assignments, c , is as follows:

- Let the state of the Markov chain consist of $c = \{c_1, \dots, c_n\}$ and $\Phi = (\phi_c; c \in c_1, \dots, c_n)$ with ϕ_c denoting cluster parameters, e.g. θ_c in our application. Repeatedly sample as follows:
- For $i = 1, \dots, n$: Let k^- be the number of distinct c_l for $l = 1, \dots, i$, and $h = k^- + m$. Label these c_l with values in $\{1, \dots, k^-\}$. If $c_i = c_l$ for some $l = 1, \dots, i$, 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 $l = 1, \dots, i$, 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 $\{1, \dots, h\}$ using the following probabilities:

$$P(c_i = c | c_{-i}, y_i, \phi_1, \dots, \phi_h) = \begin{cases} \frac{n_{-i,c}}{n-1+\alpha} F(y_i, \phi_c) & \text{for } 1 \leq c \leq k^- \\ \frac{(\alpha/m)}{n-1+\alpha} F(y_i, \phi_c) & \text{for } k^- < c \leq h \end{cases},$$

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} .

- Where $n_{-i,c}$ is the number of c_l for $l = 1, \dots, i$ 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.
- For all $c \in \{c_1, \dots, c_n\}$: Draw a new values from $\phi_c | y_i$ such that $c_i = c$, or perform some other update to ϕ_c that leaves this distribution invariant. [19]

To illustrate, $F(y_i, \phi_c)$ and $P(\phi_c | y_i)$ in our application are $F(s_{ij}, \theta_c)$ for θ_{ij} and $P(\theta_c | s_{ij}, c, \cdot)$, respectively, with:

$$F(s_{ij}, \theta_c) = \exp \left\{ - \int_B \left(\exp \left\{ a_{ij} \exp \left[- \frac{\|s_{ij} - p_i - \theta_c\|^2}{\rho} \right] \right\} \right) ds_{ij} + a_{ij} \exp \left[- \frac{\|s_{ij} - p_i - \theta_c\|^2}{\rho} \right] \right\},$$

(1)

and

$$\begin{aligned}
P(\theta_c | s_{ij}, c, \cdot) &\propto \exp \left\{ -\frac{1}{n} \sum_B \left(\exp \left\{ a_{ij} \times \exp \left[-\frac{\|s_{ij} - p_i - \theta_c\|^2}{\rho} \right] \right\} \right) \right. \\
&+ \left. \sum_{s_{ij} \in c} a_{ij} \exp \left[-\frac{\|s_{ij} - p_i - \theta_c\|^2}{\rho} \right] \right\} \\
&\times \exp \left(-\frac{1}{2} (\theta_c - \mathbf{c}_0)' \Sigma_2^{-1} (\theta_c - \mathbf{c}_0) \right), \tag{2}
\end{aligned}$$

where θ_c is the center for individual foci cluster c .

Similar expressions as in (1) and (2) are applied to p_i . To accommodate the 3-D nature of our data and to improve sampling efficiency, we modified algorithm 8 by introducing auxiliary parameters into one, two, or three dimensions of the centers at the current MCMC iteration.

Specifically, we allowed $m=7$ auxiliary parameters. For example, let $\theta_{c_i} = \{\phi_{c_i}^{(x)}, \phi_{c_i}^{(y)}, \phi_{c_i}^{(z)}\}$ be the current center for focus i assigned to cluster c . The 7 auxiliary parameters are every combination of the current cluster center and a newly generated center $\phi_0 = \{\phi_0^{(x)}, \phi_0^{(y)}, \phi_0^{(z)}\}$ from G_{02} :

- $(\phi_{c_i}^{(x)}, \phi_{c_i}^{(y)}, \phi_0^{(z)})$ updating z-dimension only
- $(\phi_{c_i}^{(x)}, \phi_0^{(y)}, \phi_{c_i}^{(z)})$ updating y-dimension only
- $(\phi_0^{(x)}, \phi_{c_i}^{(y)}, \phi_{c_i}^{(z)})$ updating x-dimension only
- $(\phi_{c_i}^{(x)}, \phi_0^{(y)}, \phi_0^{(z)})$ updating y and z-dimensions only
- $(\phi_0^{(x)}, \phi_{c_i}^{(y)}, \phi_0^{(z)})$ updating x and y-dimensions only
- $(\phi_0^{(x)}, \phi_0^{(y)}, \phi_{c_i}^{(z)})$ updating x and y-dimensions only
- $(\phi_0^{(x)}, \phi_0^{(y)}, \phi_0^{(z)})$ updating all dimensions only.

By designing the auxiliary parameters as above, i.e., based on the current sampled values, the probability of making a movement is improved compared to randomly generating the auxiliary parameters from the base distribution. To further speed up the convergence, we limit the choice of auxiliary parameters within the defined brain space. All these efforts are aimed to improve MCMC sampling efficiency and achieve quick convergence.

2.4 Determining the clusters

To estimate cluster centers and infer cluster assignment, we implement the least-squared Euclidean distance method introduced in [20], and outline below.

1. After the MCMC burn-in, continue the MCMC simulations for an additional W iterations. Let A denote an $n \times n$ matrix. The $(i, j)^{th}$ entry of A is the proportion of iterations such that foci i and j ($i, j = 1, \dots, n$) are in the same cluster. The matrix A is referred to as an averaged clustering matrix.

2. Continue to run an additional F iterations of the MCMC simulations. For each iteration,
 - a. form an $n \times n$ matrix composed of indicators of clustering for that particular iteration. For instance, if foci i and j are in one cluster, then the $(i, j)^{th}$ entry is 1; otherwise, it is zero.
 - b. calculate the Euclidean distance between the matrix formed above and the averaged clustering matrix A .
3. Sort the Euclidean distances obtained from the F iterations, and the final selection on the number of clusters is in favor of simpler clusters and relatively small Euclidean distances. The parameters will then be inferred accordingly based on the identified clusters.

Determining the clusters in this way enable us to incorporate information collected in all the iterations after burn-in [20]. The cluster patterns to be inferred and summarized will focus on individual foci clusters as they are our primary interest.

2.5 Selection of α

The specification of α can have a potentially significant effect on the number of clusters identified due to its direct impact on the aggregation of G about G_0 . A smaller choice of α places less weight upon the base distribution, resulting in a smaller number of clusters. The extent of this sensitivity and various ways to estimate α have been discussed rigorously in the literature [21], [22], [23], but so far an objective and efficient method for determining α is not available. Instead, given the importance of α , we decide to select α based on information in the data, i.e., posterior joint likelihood of the parameters. Specifically, the choice of α_1 for p_i and α_2 for θ_{ij} were selected iteratively based on a grid search on a set of possible values for α_1 and α_2 that minimize and stabilize the deviance information criterion (DIC) [24].

3 Inference

3.1 Simulation designs

Simulations were performed to demonstrate and assess the proposed method. In total, we considered 50 studies each with 10 foci or sub-studies. Three individual foci clusters are simulated centered at (1,1,1), (2,2,2), and (4,4,4) covering 150, 150, and 200 foci, respectively. Two study centers are assumed with centers held at (0.1, 0.1, 0.1) and (0.4, 0.4, 0.4) with each including 25 studies. We considered the following four scenarios:

1. The first scenario is used to demonstrate the proposed method. We simulate data via multivariate normal with mean at the individual foci centers and variance small $\Sigma = 0.002I_3$ (i.e., the standard deviation is 0.045). In the fourth scenario below, we assess the robustness of the method with respect to variations in data.
2. The second scenario is to demonstrate its ability to cluster outliers. We follow the same setting as in scenario 1) but added an additional focus in the third

individual foci cluster located in the right 5% tail of a multivariate normal with mean (4, 4, 4) and variance as in scenario 1.

3. It is important to examine the robustness of the method with respect to abnormal patterns. We follow the same scenarios as in 1) to simulate clusters 1 and 2. Cluster 3 is simulated using truncated normal distribution with mean $\alpha(4, 4, 4)$ and variance $0.002I_3$. The lower bound is set at $\alpha(1, 1, 1)$ and upper bound at ∞ .
4. It is often that the boundaries between clusters are not clear due to large variance. The last scenario is aimed to assess the sensitivity of the method when the variance increases. Besides $\Sigma = 0.002I_3$, we considered additional four levels of Σ : $\Sigma = 0.01I_3, 0.05I_3, 0.1I_3, \text{ and } 0.2I_3$. Other settings are as in scenario 1).

In total, 100 Monte Carlo (MC) replicates are generated under each scenario. For each setting, the values of α_1 and α_2 are determined using the method discussed in Section 2.5 based on one randomly selected MC replicate. In total, 3,500 converged MCMC iterations are used for this purpose. After α_1 and α_2 are determined, for each data set, we used 1,500 as burn in iterations, 2,500 iterations to determine the average clustering matrix, and 1,000 additional iterations to infer the clusters and individual foci cluster centers.

Three statistics are used to assess the method: sensitivity, specificity, and percentage of correct clustering. Sensitivity is defined per cluster as the proportion of foci that are correctly assigned to that given cluster, $Se=TP/(TP+FN)$ and specificity is defined per cluster as the proportion of foci that are correctly not assigned to a cluster, $Sp=TN/(TN+FP)$. In these definitions, true positive (TP) denotes that a focus in a respective cluster is also assigned to that cluster, false negative (FN) denotes that a focus is in a respective cluster but not assigned to that cluster, true negative (TN) is a focus that is not in the respective cluster and not assigned to that cluster, and false positive (FP) denotes a focus that is not in the respective cluster but assigned to that cluster. Percentage of correct clustering is an overall measure defined as the proportion of foci that are correctly clustered. Note that the definition of correctness takes into account both TP and TN. The program to implement the method was written in R and the authors will gladly provide the programs to researchers upon request.

3.2 Results

To illustrate the choice of α_1 and α_2 via grid search, we use data generated from scenario 1). As indicated in Figure 1, DIC is stabilized in all situations, and the optimal choice for α_1 and α_2 is $\alpha_1 = 0.1$ and $\alpha_2 = 3$. These two precision parameters are then utilized in all simulated data generated under scenario 1) to infer the individual foci clusters. Our further simulations indicate that neighboring values of the selected α_1 and α_2 in general give consistent findings.

Table 1 summaries the findings on individual foci cluster identification and the quality of the identified clusters. Overall, the method is robust with respect to outliers, skewness, and large variation in each foci clusters. Among the 100 MC replicates, the proposed method correctly identified clusters in at least 89% of the replicates across all scenarios except for the situation of the largest variance ($0.2I_3$), in which case the percentage of correct identification

is still high enough, 80%. As variances increase, the sensitivity and correctness decrease. This is expected given the potential that two clusters may overlap as variances get large. Unlike variances, outliers and skew foci cluster patterns do not impact the quality of identified foci clusters; sensitivity, specificity, and percentage of correct foci cluster identification are all high (Table 1).

The scenarios we chose represent important facets of variability that a spatial models need to be able to handle and overcome in order to accurately perform. Based on our simulations, it can be inferred that across all the scenarios, our proposed model in general performed well in correctly identifying individual foci clusters.

3.3 Further simulations for comparison

To further assess the proposed method, we designed additional simulation scenarios to examine how sensitive the method is with respect to the choice of prior distributions of variance parameters and the choice of abnormal distributions of foci, as well as to compare with existing methods for clustering such as the K-means approach.

Scenarios—For each of the following scenarios, we simulate 10 MC replicates and, as before, the statistics to evaluate the quality of identified clusters include means and standard deviations of sensitivity, specificity, and correctness rate.

1. Sensitivity to the choice of prior distribution of variance parameters in the base distributions. In the earlier setting, we took $IG(0.5, 0.5)$ as the prior distribution following the suggestion by Kass and Wasserman [18]. We considered an additional prior distribution, $IG(0.001, 0.001)$, a vague prior commonly chosen for variance parameters. Each MC replicates is generated based on the first simulation scenario discussed previously.
2. Sensitivity to skewness pattern of foci distribution. The truncated multivariate normal distribution utilized to simulate foci with skew pattern (the third simulation scenario) has a short tail and is still closely related to normal distribution. To generate data with much longer tail, we revised simulation scenario 3 by using a chi-square distribution with mean of 4 and a degrees of freedom being 0.5 for each dimension. Other settings were remained the same as in the original scenario 3.
3. Comparison with existing clustering methods. The proposed method has the potential to cluster foci abnormally distributed and has the ability to adjust other factor effects, e.g., study effect, when clustering individual foci. We expect it outperforms existing clustering methods; here, we use the commonly used K-means approach as a representative. To perform the comparison, each MC replicate is simulated based on the first simulation scenario with study effects centered at 1 and 2. In the clustering process, for the proposed method, prior knowledge of study effects is incorporated into the clustering process.

Results—Table 2 summarizes the results for these further simulation scenarios. The findings demonstrate that the method is reasonably insensitive to the choice of prior

distributions for variance parameters in the Dirichlet process, as seen by the statistics of sensitivity, specificity, and accuracy, which are comparable to the results when using $IG(0.5, 0.5)$ as the prior distribution. Nevertheless, it is worth noting that posterior inferences can be influenced by the choice of prior distributions, especially when the priors are informative [25]. When data generated from a chi-square distribution, although the data are severely right skewed, the proposed method is still able to correctly cluster individual foci 90% of the time. Lastly, when comparing our method with the K-means approach, our method showed much higher sensitivity, specificity, and accuracy, compared to the K-means approach, which is lack of the ability of incorporating prior knowledge into the clustering process.

4 Data Application

We applied the proposed model to the meta analysis data set implemented in [8] and [5] (Figure 2). This meta data were constructed from 164 different neuroimaging publications. Each publication had a number of statistical comparisons or contrasts within it, totaling 437 contrasts. We identify these contrasts as “studies”. In terms of individual foci, 2,478 were reported ([5], [8]). As seen in Table 3, on average there were 15.11 foci per publication and 5.67 foci per study. Additionally there was an average of 2.67 studies per publication. Affective (Table 4) was the most frequent emotion found in 175 studies and surprise the least frequent emotion found in only 2 studies.

Clustering across anatomical structures or other specific regions of interest are important issues in the neuroscience field. Therefore, the Automated Anatomical Labeling [26] system was utilized and identified 116 brain regions of interest (ROI). Not all foci fell within these regions of interest, 498 did not fall within a ROI; however, they were included within the analysis due to their potential influence on cluster centers.

We were advised by researchers familiar with the data construction that the study effects were likely to be quite small. Introducing this prior knowledge into the application of the prior distributions for μ (the mean in the base distribution for study effect) and for σ_1^2 , we set $\sigma_\mu^2=0.04$ (standard deviation is 0.063) and took $\sigma_1^2 \sim IG(3, 0.5)$. Accordingly, the amount of truncation for the base distribution in modeling study effects was set at 10% ($r_a = r_b = 10\%$). Other prior assumptions were as those proposed in the simulations. To avoid computational overflow, the data was re-scaled down by 10.

4.1 Results

As in the simulations, α_1 and α_2 were determined by use of grid search, which resulted in $\alpha_1 = 0.1$ and $\alpha_2 = 3$. We then ran 8,000 iterations with 4,000 for burn-in. Trace plots were used to monitor the convergence of the MCMC chains. To illustrate, trace plots for posterior samples of σ_1^2 and σ_2^2 were included in Figure 3, showing reasonable convergence of the chains after 4,000 iterations.

The proposed method identified 46 study clusters and 44 individual clusters, of which 34 individual clusters were located in a ROI (Table 5) and were included in our following discussions. The break down of each cluster by its foci frequency and emotion frequency can

be seen in Table 6. The largest cluster is cluster 1 centered at (26.89, -0.2, -10.10). It was located in the right amygdala region of the brain and contained foci from all 8 emotions. The next two largest clusters identified were clusters 2 and 5. Cluster 2 had 122 foci; it was related to 7 emotions and located in the left frontal inferior triangularis. Cluster 5 (101 foci) was located in the left amygdala and also related to 7 emotions.

We identified cluster centers located in 18 different regions of the brain, 25 if differentiating between left and right hemispheres. The functionality of the majority of these regions have been associated with emotions or emotional response, task performance, or consciousness/self-awareness. For example, two of the largest clusters are located in the amygdala, a region typically activated when subjects are emotional provoked [27]. These regions are also identified by studies elsewhere as areas of high activation [8]. Other regions that have been found associated with emotions due to brain activation are cingulum, insula, thalamus, and inferior frontal orbitofrontal [27] and these regions are also included in the clusters identified by the proposed method. Given the agreement of the findings with the literature, the clustering results have the potential to inform the centers of those activation regions.

We next compared the results with those from the K-means approach. By maximizing between cluster variations and minimizing the within cluster variation, the K-means approach identified 20 clusters, centered in 19 different brain regions, 20 if hemisphere is differentiated. With the exception of the cerebellum 6 (left and right), cerebellum 4 5, and the rolandic operculum region, all brain regions were consistent with those identified by the proposed method. The two largest clusters identified by the K-means approach are located in the amygdala and hippocampus, respectively. Given the small study effects, the proposed method and the K-means approach showed certain agreement in this application. However, as seen in Figure 2, the foci tends to be skew distributed. Many existing clustering methods including the K-means approach are sensitive to non-spherical patterns, which may explain the difficulty of the K-means approach in differentiating between different clusters, and consequently the result of smaller number of clusters.

5 Discussion

The proposed spatial Cox point process model with a Gaussian kernel driven intensity function was motivated by the need to spatially cluster coordinate-based meta analysis data to identify activated regions with the brain. The Gaussian kernel incorporated a study and sub-study effects that were estimated using a Dirichlet process (DP). The advantage of the proposed method exists in its ability to adjust for study effect (or other factor effects) and to fit irregular spatial patterns.

Findings from simulations indicated that the method is robust with respect to outliers and skewed distributions. For clusters with wide variations, the method still performed well enough as long as the overlaps between clusters are not substantial. The model was further applied to an emotion meta data set. In total, 34 clusters located in regions of interest were identified, and 14 of the 34 were in the regions known to be associated with emotion. Several regions of interest appear in multiple clusters suggest potential overlapping over clusters. It also stands to mention the natural limitation of this meta-data as it contains

studies from both fMRI and positron emission tomography (PET) scans. FMRI and PET scans have different resolutions and time constraints. Thus, brain activations identified in one may not be consistent in the other.

Multiple future research directions exist. First, when the study effect is large and we do not have any prior knowledge regarding its strength, the estimates on the number of clusters will be unbiased but the estimates of the centers can be biased. This issue of identifiability can be solved by constructing parameter constraints in the clustering process, but given the dynamic feature of the clustering process, a great effort is needed to ensure that the same constraint is implemented from one MCMC iteration to the next. Additionally, due to the nature of the Dirichlet process for estimating distribution functions, using DP to perform cluster analysis can possibly produce redundant clusters, e.g., two clusters identified but their centers are so close that they should be viewed as one cluster. There is a need to design an approach which has the ability to merge clusters with centers close to each other. One possible direction is to consider mixtures of Dirichlet process to describe foci distribution patterns. In this case, two groups of foci with close centers will likely be included in one cluster. However, by use of mixtures of Dirichlet process, the modeling will be more complicated compared to the proposed method and techniques to simplify the sampling process are necessary.

Acknowledgments

The authors would like to thank Professor Tor D. Wager for providing the meta-analysis data. Dr. Kang's effort was supported by an NIH grant 1R01MH105561. Dr. Zhang and Dr. Ray's efforts were supported by their start funds provided by the School of Public Health at the University of Memphis.

References

1. Logothetis NK. What we can do and what we cannot do with fmri. *Nature*. 2008; 453(7197):869–878. [PubMed: 18548064]
2. Smith SM. Overview of fmri analysis. *British journal of radiology*. 2004; 77(suppl 2):S167–S175. [PubMed: 15677358]
3. Thirion B, Pinel P, Mériaux S, Roche A, Dehaene S, Poline JB. Analysis of a large fmri cohort: Statistical and methodological issues for group analyses. *Neuroimage*. 2007; 35(1):105–120. [PubMed: 17239619]
4. Fox PT, Lancaster JL, Parsons LM, Xiong JH, Zamarripa F. Functional volumes modeling: Theory and preliminary assessment. *Human Brain Mapping*. 1997; 5(4):306–311. [PubMed: 20408233]
5. Kober H, Barrett LF, Joseph J, Bliss-Moreau E, Lindquist K, Wager TD. Functional grouping and cortical–subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage*. 2008; 42(2):998–1031. [PubMed: 18579414]
6. Eickhoff SB, Laird AR, Grefkes C, Wang LE, Zilles K, Fox PT. Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: A random-effects approach based on empirical estimates of spatial uncertainty. *Human brain mapping*. 2009; 30(9):2907–2926. [PubMed: 19172646]
7. Salimi-Khorshidi G, Smith SM, Keltner JR, Wager TD, Nichols TE. Meta-analysis of neuroimaging data: a comparison of image-based and coordinate-based pooling of studies. *Neuroimage*. 2009; 45(3):810–823. [PubMed: 19166944]
8. Kang J, Johnson TD, Nichols TE, Wager TD. Meta analysis of functional neuroimaging data via bayesian spatial point processes. *Journal of the American Statistical Association*. 2011; 106(493): 124–134. [PubMed: 21706069]

9. Turkeltaub PE, Eden GF, Jones KM, Zeffiro TA. Meta-analysis of the functional neuroanatomy of single-word reading: method and validation. *Neuroimage*. 2002; 16(3):765–780. [PubMed: 12169260]
10. Laird AR, Fox PM, Price CJ, Glahn DC, Uecker AM, Lancaster JL, Turkeltaub PE, Kochunov P, Fox PT. Ale meta-analysis: Controlling the false discovery rate and performing statistical contrasts. *Human brain mapping*. 2005; 25(1):155–164. [PubMed: 15846811]
11. Turkeltaub PE, Eickhoff SB, Laird AR, Fox M, Wiener M, Fox P. Minimizing within-experiment and within-group effects in activation likelihood estimation meta-analyses. *Human brain mapping*. 2012; 33(1):1–13. [PubMed: 21305667]
12. Eickhoff SB, Bzdok D, Laird AR, Kurth F, Fox PT. Activation likelihood estimation meta-analysis revisited. *Neuroimage*. 2012; 59(3):2349–2361. [PubMed: 21963913]
13. Wager TD, Jonides J, Reading S. Neuroimaging studies of shifting attention: a meta-analysis. *Neuroimage*. 2004; 22(4):1679–1693. [PubMed: 15275924]
14. Wager TD, Lindquist M, Kaplan L. Meta-analysis of functional neuroimaging data: current and future directions. *Social cognitive and affective neuroscience*. 2007; 2(2):150–158. [PubMed: 18985131]
15. Wager TD, Phan KL, Liberzon I, Taylor SF. Valence, gender, and lateralization of functional brain anatomy in emotion: a meta-analysis of findings from neuroimaging. *Neuroimage*. 2003; 19(3): 513–531. [PubMed: 12880784]
16. Moller, J., Waagepetersen, RP. *Statistical inference and simulation for spatial point processes*. CRC Press; 2004.
17. Escobar MD, West M. Bayesian density estimation and inference using mixtures. *Journal of the american statistical association*. 1995; 90(430):577–588.
18. Kass RE, Wasserman L. A reference bayesian test for nested hypotheses and its relationship to the schwarz criterion. *Journal of the American Statistical Association*. 1995; 90(431):928–934.
19. Neal RM. Markov chain sampling methods for dirichlet process mixture models. *Journal of computational and graphical statistics*. 2000; 9(2):249–265.
20. Dahl D. Model-based clustering for expression data via a dirichlet process mixture model. *Bayesian inference for gene expression and proteomics*. 2006:201–218.
21. Liu JS. Nonparametric hierarchical bayes via sequential imputations. *The Annals of Statistics*. 1996; 24(3):911–930.
22. McAuliffe JD, Blei DM, Jordan MI. Nonparametric empirical bayes for the dirichlet process mixture model. *Statistics and Computing*. 2006; 16(1):5–14.
23. Kyung M, Gill J, Casella G. Estimation in dirichlet random effects models. *The Annals of Statistics*. 2010; 38(2):979–1009.
24. Congdon, P. *Bayesian statistical modelling*. Vol. 704. John Wiley & Sons; 2007.
25. Gelman, A., Carlin, JB., Stern, HS., Rubin, DB. *Bayesian data analysis*. Vol. 2. Taylor & Francis; 2014.
26. Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M. Automated anatomical labeling of activations in spm using a macroscopic anatomical parcellation of the mni mri single-subject brain. *Neuroimage*. 2002; 15(1):273–289. [PubMed: 11771995]
27. Keightley ML, Winocur G, Graham SJ, Mayberg HS, Hevenor SJ, Grady CL. An fmri study investigating cognitive modulation of brain regions associated with emotional processing of visual stimuli. *Neuropsychologia*. 2003; 41(5):585–596. [PubMed: 12559151]

Biographies

Meredith Ray is serving as an assistant professor in biostatistics at the University of Memphis for the School of Public Health in the Division of Epidemiology, Biostatistics, and Environment health. Her research focuses on Bayesian methods, semi-parametric modeling, and fMRI brain imaging data analysis.

Jian Kang is an Assistant Professor of Biostatistics at the University of Michigan. He is also affiliated with Kidney Epidemiology and Cost Center, at the University of Michigan. His research focuses on statistical methods and theory for large-scale complex biomedical data. He is particularly interested in Bayesian methods, high dimensional variable selection with applications to imaging and genetic studies.

Hongmei Zhang is serving as an associate professor in biostatistics at the University of Memphis for the School of Public Health in the Division of Epidemiology, Biostatistics, and Environment health. Her research focuses on Bayesian methods, bioinformatics, variable selection, clustering methods, and measurement error modeling.

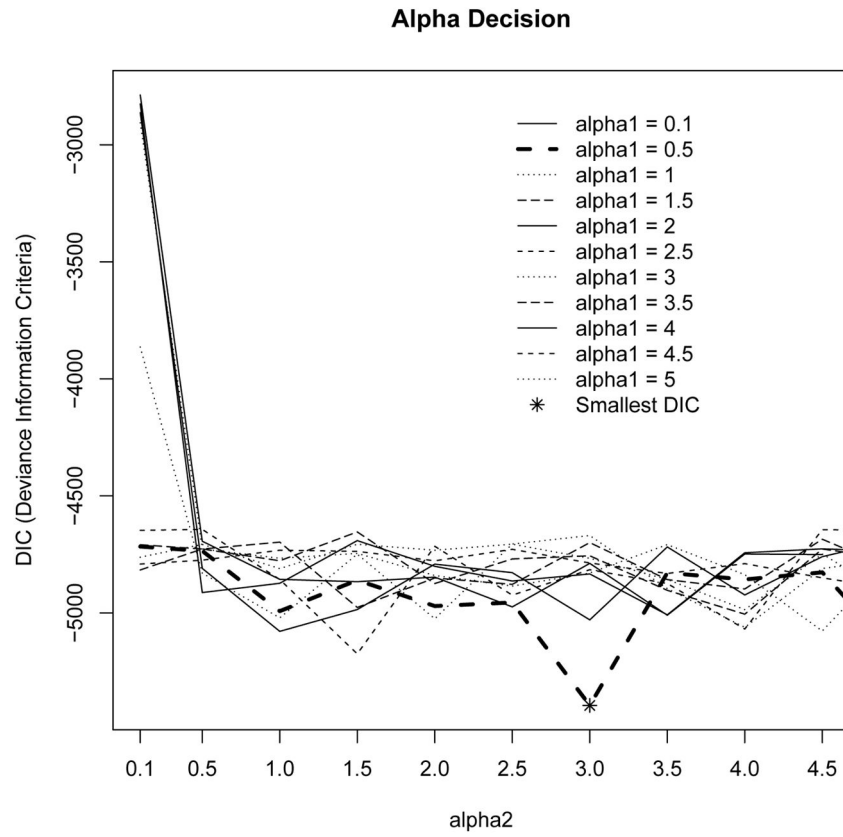


Fig. 1. Example of grid search for selection of α_1 and α_2 based on deviance information criterion.

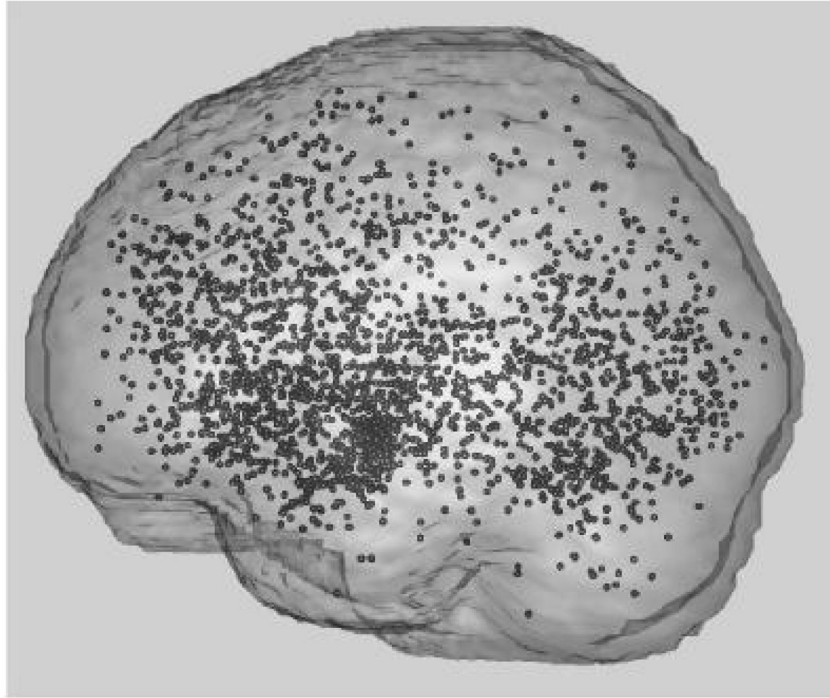


Fig. 2.
The fMRI meta-data implemented in Kang et al. [8]

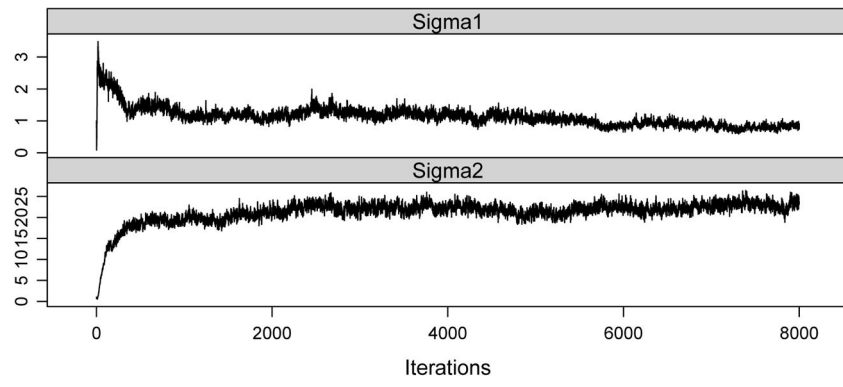


Fig. 3.
Convergence plots for the conditional sampling of σ_1^2 and σ_2^2 .

Summary Statistics of Clustering Results under Different Simulation Scenarios (Simulated Data).

TABLE 1

Scenario:	Median Num. of Clusters (SD) *	Cluster index	Average Sensitivity (SD) *	Average Specificity (SD) *	Average % Correctness rate (SD) *
Normal	** IC: 4 (0.9)	1	0.99 (0.01)	1.00 (0.00)	99.00 (1.00)
		2	0.99 (0.01)	1.00 (0.00)	99.00 (1.00)
		3	1.00 (0.00)	1.00 (0.00)	99.00 (1.00)
Outlier	** SC: 2 (0.9)	1	0.98 (0.04)	1.00 (0.00)	98.00 (5.00)
		2	0.97 (0.09)	1.00 (0.01)	98.00 (5.00)
Skewed	IC: 4 (0.82)	1	0.99 (0.00)	1.00 (0.00)	100.00 (0.00)
		2	0.99 (0.01)	1.00 (0.00)	100.00 (0.00)
		3	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
Skewed	** SC: 2 (0.82)	1	0.95 (0.12)	1.00 (0.00)	95.00 (7.00)
		2	0.96 (0.10)	1.00 (0.01)	95.00 (7.00)
Large1	SC: 4.5 (0.42)	1	0.85 (0.18)	1.00 (0.00)	85.00 (12.00)
		2	0.85 (0.16)	1.00 (0.00)	85.00 (12.00)
		3	0.85 (0.31)	1.00 (0.00)	85.00 (12.00)
Large1	IC: 3 (0.41)	1	0.92 (0.27)	0.96 (0.11)	90.00 (19.00)
		2	0.94 (0.24)	0.96 (0.10)	90.00 (19.00)
		3	0.85 (0.31)	0.93 (0.17)	90.00 (19.00)
Large1	SC: 4 (0.41)	1	0.84 (0.18)	1.00 (0.02)	85.00 (12.00)
		2	0.85 (0.16)	1.00 (0.02)	85.00 (12.00)
Large2	IC: 3 (0.45)	1	0.99 (0.01)	1.00 (0.00)	100.00 (0.00)
		2	0.99 (0.01)	1.00 (0.00)	100.00 (0.00)
		3	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
Large2	SC: 2 (0.45)	1	0.99 (0.02)	0.99 (0.02)	99.00 (2.00)

Scenario:	Median Num. of Clusters (SD) *	Cluster Index	Average Sensitivity (SD) *	Average Specificity (SD) *	Average % Correctness rate (SD) *
		2	0.99 (0.03)	1.00 (0.01)	
Large3		1	0.97 (0.1)	0.93 (0.12)	
	IC: 3 (0.44)	2	0.96 (0.14)	0.93 (0.11)	89.00 (17.00)
		3	0.78 (0.4)	0.98 (0.08)	
Large3	SC: 4 (0.44)	1	0.84 (0.17)	0.97 (0.05)	85.00 (11.00)
		2	0.87 (0.14)	0.96 (0.05)	
Large4		1	0.72 (0.18)	0.97 (0.01)	
	IC: 9 (2.13)	2	0.71 (0.19)	0.97 (0.01)	80.00 (8.00)
		3	0.93 (0.07)	0.98 (0.01)	
Large4	SC: 4 (2.13)	1	0.71 (0.21)	0.91 (0.09)	73.00 (15.00)
		2	0.75 (0.17)	0.88 (0.12)	

* SD: standard deviation across 100 MC replicates;

** IC: individual foci cluster; SC: study effect clusters; Covariance matrices: $\Sigma = 0.01I_3$ (Large1), $0.05I_3$ (Large2), $0.1I_3$ (Large3), and $0.2I_3$ (Large4).

Assessment of the Robustness of the Method and a Comparison with the K-means Approach (Simulated Data).

TABLE 2

Scenario:	Median Num. of Clusters (SD) *	Cluster index	Average Sensitivity (SD) *	Average Specificity (SD) *	Average % Correctness rate (SD) *
Vague Prior	** IC: 3 (0.32)	1	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
		2	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
		3	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
** SC: 2.5 (0.32)	1	0.99 (0.02)	1.00 (0.00)	97.00 (4.00)	
	2	0.96 (0.07)	1.00 (0.00)	97.00 (4.00)	
Skewed with Chi-Squared	IC: 5.5 (2.42)	1	1.00 (0.00)	0.96 (0.01)	90.00 (2.00)
		2	0.99 (0.01)	0.95 (0.01)	90.00 (2.00)
		3	0.76 (0.06)	1.00 (0.00)	90.00 (2.00)
SC: 4 (2.41)	1	0.78 (0.17)	1.00 (0.00)	81.00 (11.00)	
	2	0.84 (0.18)	1.00 (0.01)	81.00 (11.00)	
Our method IC: 3 (0.63)		1	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
		2	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
		3	1.00 (0.00)	1.00 (0.00)	100.00 (0.00)
K-means vs Our method		1	0.55 (0.5)	0.89 (0.11)	69.00 (19.00)
		2	0.55 (0.28)	0.83 (0.2)	69.00 (19.00)
		3	0.90 (0.21)	0.98 (0.08)	69.00 (19.00)

* SD: standard deviation across 10 MC replicates;

** IC: individual foci cluster; SC: study effect clusters.

TABLE 3

Descriptive Statistics of the Meta Data.

	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 4

Frequency of Emotions (Meta Data).

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%)
happy	36 (8.24%)	178 (7.18%)
mixed	41 (9.38%)	195 (7.87%)
sad	45 (10.3%)	348 (14.04%)
surprise	2 (0.46%)	6 (0.24%)
Total	437	2478

* aff: affective.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

TABLE 5

Meta Data Cluster Results Based on the Proposed Method.

Individual Foci Clusters						
Cluster Centers	Brain Regions	Cluster Index	# of foci per cluster (% of total foci)	# of foci not in ROI (% of cluster foci)	# of studies per cluster (% of all studies)	
(26.89,-0.2,-10.1)	Amygdala R	1	227 (9.16)	51 (22.47)	169 (38.67)	
(-38.93,30.28,-0.54)	Frontal Inf Tri L	2	122 (4.92)	18 (14.75)	99 (22.65)	
(-15.51,-0.42,-11.88)	Amygdala L	3	101 (4.08)	25 (24.75)	84 (19.22)	
(48.47,12.21,-13.08)	Temporal Pole Sup R	4	91 (3.67)	21 (23.08)	81 (18.54)	
(53.66,29.64,14.95)	Frontal Inf Tri R	5	91 (3.67)	14 (15.38)	76 (17.39)	
(2.05,32.23,32.04)	Cingulum Ant L	6	90 (3.63)	21 (23.33)	75 (17.16)	
(5.83,40.37,4.93)	Cingulum Ant R	7	81 (3.27)	15 (18.52)	67 (15.33)	
(-32.4,-63.7,-13.15)	Fusiform L	8	79 (3.19)	25 (31.65)	66 (15.1)	
(19.16,-14.05,3.23)	Thalamus R	9	78 (3.15)	36 (46.15)	66 (15.1)	
(52.37,-61.45,5.89)	Temporal Mid R	10	72 (2.91)	10 (13.89)	65 (14.87)	
(45.04,-46.63,-17.55)	Fusiform R	11	69 (2.78)	14 (20.29)	64 (14.65)	
(-3.09,-87.92,0.41)	Calcarine L	12	69 (2.78)	10 (14.49)	64 (14.65)	
(-26.05,13.34,-15)	Insula L	13	67 (2.7)	15 (22.39)	60 (13.73)	
(-33.05,-1.88,0.76)	Insula L	14	63 (2.54)	15 (23.81)	56 (12.81)	
(-36.93,2.86,8.66)	Insula L	15	59 (2.38)	15 (25.42)	53 (12.13)	
(24.44,-49.21,-11.97)	Fusiform R	16	58 (2.34)	10 (17.24)	52 (11.9)	
(4.56,53.78,37.04)	Frontal Sup Medial L	17	57 (2.3)	3 (5.26)	50 (11.44)	
(57.35,-40.14,11.52)	Temporal Mid R	18	55 (2.22)	11 (20)	48 (10.98)	
(-25.2,34.18,-6.25)	Frontal Inf Orb L	19	53 (2.14)	4 (7.55)	50 (11.44)	
(3.6,-58.78,24.09)	Precuneus R	20	53 (2.14)	8 (15.09)	46 (10.53)	
(-36.05,-49.14,-14.47)	Fusiform L	21	44 (1.78)	5 (11.36)	38 (8.7)	
(35.74,33.65,-4.05)	Frontal Inf Orb R	22	44 (1.78)	8 (18.18)	43 (9.84)	
(46.33,33.4,-6.22)	Frontal Inf Orb R	23	44 (1.78)	9 (20.45)	42 (9.61)	
(47.43,19.45,5.17)	Frontal Inf Tri R	24	43 (1.74)	5 (11.63)	37 (8.47)	
(44.35,2.6,12.7)	Insula R	25	41 (1.65)	9 (21.95)	39 (8.92)	
(-35.76,9.07,36.63)	Frontal Mid L	26	41 (1.65)	6 (14.63)	39 (8.92)	
(6.52,10.28,64.77)	Supp Motor Area R	27	34 (1.37)	4 (11.76)	30 (6.86)	

Individual Foci Clusters						
Cluster Centers	Brain Regions	Cluster Index	# of foci per cluster (% of total foci)	# of foci not in ROI (% of cluster foci)	# of studies per cluster (% of all studies)	
(-39.51,-68.82,15.92)	Temporal Mid L	28	25 (1.01)	4 (16)	25 (5.72)	
(30.72,9.29,7.29)	Putamen R	29	21 (0.85)	3 (14.29)	21 (4.81)	
(45.86,-46.89,27.05)	Angular R	30	8 (0.32)	1 (12.5)	8 (1.83)	
(-33.37,-1.77,47.85)	Precentral L	31	7 (0.28)	1 (14.29)	7 (1.6)	
(1.15,-75.77,14.66)	Calcarine L	32	7 (0.28)	0 (0)	7 (1.6)	
(46.48,19.42,0.33)	Insula R	33	5 (0.2)	0 (0)	5 (1.14)	
(5.92,-3.57,43.04)	Cingulum Mid R	34	5 (0.2)	2 (40)	5 (1.14)	

R: right hemisphere, L: left hemisphere.

TABLE 6

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	227	Cluster: 2	122	Cluster: 3	101	Cluster: 4	91	Cluster: 5	91	Cluster: 6	90	Cluster: 7	81	Cluster: 8	79	Cluster: 9	78	Cluster: 10	72
aff	86 (37.89)	aff	42 (34.43)	aff	27 (26.73)	aff	36 (39.56)	aff	30 (32.97)										
fear	40 (17.62)	disgust	15 (12.3)	fear	23 (22.77)	anger	12 (13.19)	fear	18 (19.78)										
disgust	32 (14.1)	sad	15 (12.3)	sad	18 (17.82)	sad	12 (13.19)	disgust	17 (18.68)										
sad	25 (11.01)	fear	14 (11.48)	disgust	12 (11.88)	fear	10 (10.99)	sad	11 (12.09)										
mixed	18 (7.93)	mixed	13 (10.66)	anger	7 (6.93)	disgust	8 (8.79)	mixed	9 (9.89)										
happy	14 (6.17)	happy	12 (9.84)	happy	7 (6.93)	happy	7 (7.69)	happy	4 (4.4)										
anger	11 (4.85)	anger	11 (9.02)	mixed	7 (6.93)	mixed	6 (6.59)	anger	2 (2.2)										
surprise	1 (0.44)																		
Cluster: 6	90	Cluster: 7	81	Cluster: 8	79	Cluster: 9	78	Cluster: 10	72										
aff	29 (32.22)	aff	32 (39.51)	aff	29 (36.71)	aff	25 (32.05)	aff	25 (34.72)										
fear	18 (20)	disgust	11 (13.58)	fear	16 (20.25)	sad	15 (19.23)	fear	16 (22.22)										
sad	12 (13.33)	fear	11 (13.58)	disgust	13 (16.46)	disgust	12 (15.38)	sad	14 (19.44)										
disgust	11 (12.22)	anger	9 (11.11)	sad	13 (16.46)	fear	8 (10.26)	disgust	8 (11.11)										
mixed	8 (8.89)	sad	8 (9.88)	anger	3 (3.8)	anger	7 (8.97)	mixed	4 (5.56)										
happy	7 (7.78)	happy	5 (6.17)	happy	3 (3.8)	mixed	6 (7.69)	happy	3 (4.17)										
anger	5 (5.56)	mixed	5 (6.17)	mixed	2 (2.53)	happy	4 (5.13)	anger	2 (2.78)										
						surprise	1 (1.28)												
Cluster: 11	69	Cluster: 12	69	Cluster: 13	67	Cluster: 14	63	Cluster: 15	59										
aff	20 (28.99)	aff	27 (39.13)	aff	29 (43.28)	aff	19 (30.16)	aff	22 (37.29)										
fear	11 (15.94)	disgust	13 (18.84)	fear	11 (16.42)	fear	11 (17.46)	disgust	11 (18.64)										
mixed	10 (14.49)	happy	9 (13.04)	sad	11 (16.42)	disgust	10 (15.87)	sad	8 (13.56)										
disgust	9 (13.04)	fear	6 (8.7)	disgust	6 (8.96)	happy	8 (12.7)	anger	5 (8.47)										
sad	9 (13.04)	mixed	6 (8.7)	happy	5 (7.46)	sad	7 (11.11)	fear	5 (8.47)										
anger	5 (7.25)	sad	5 (7.25)	anger	3 (4.48)	anger	5 (7.94)	happy	4 (6.78)										

Cluster Index: Total foci in that cluster											
Emotion Frequency of emotion (% of total cluster foci)											
happy	5 (7.25)	anger	3 (4.35)	mixed	2 (2.99)	mixed	3 (4.76)	mixed	4 (6.78)	Cluster: 16	53
aff	23 (39.66)	aff	18 (31.58)	aff	17 (30.91)	aff	22 (41.51)	aff	20 (37.74)	Cluster: 17	53
fear	9 (15.52)	sad	12 (21.05)	disgust	9 (16.36)	fear	10 (18.87)	fear	7 (13.21)	Cluster: 18	55
disgust	8 (13.79)	fear	8 (14.04)	mixed	8 (14.55)	sad	8 (15.09)	mixed	7 (13.21)	Cluster: 19	53
sad	7 (12.07)	anger	6 (10.53)	fear	7 (12.73)	disgust	6 (11.32)	disgust	6 (11.32)	Cluster: 20	53
mixed	6 (10.34)	happy	6 (10.53)	anger	6 (10.91)	mixed	4 (7.55)	happy	5 (9.43)	Cluster: 21	44
anger	3 (5.17)	disgust	4 (7.02)	sad	4 (7.27)	happy	2 (3.77)	sad	5 (9.43)	Cluster: 22	44
happy	2 (3.45)	mixed	3 (5.26)	happy	3 (5.45)	anger	1 (1.89)	anger	3 (5.66)	Cluster: 23	44
		surprise	1 (1.82)	surprise	1 (1.82)					Cluster: 24	43
										Cluster: 25	41
aff	19 (43.18)	aff	19 (43.18)	aff	17 (38.64)	aff	15 (34.88)	aff	21 (51.22)	Cluster: 26	41
sad	7 (15.91)	disgust	7 (15.91)	disgust	8 (18.18)	disgust	7 (16.28)	sad	8 (19.51)	Cluster: 27	34
fear	6 (13.64)	sad	7 (15.91)	mixed	8 (18.18)	fear	7 (16.28)	disgust	3 (7.32)	Cluster: 28	34
disgust	5 (11.36)	mixed	4 (9.09)	fear	5 (11.36)	sad	7 (16.28)	fear	3 (7.32)	Cluster: 29	25
anger	3 (6.82)	fear	3 (6.82)	happy	3 (6.82)	anger	4 (9.3)	anger	2 (4.88)	Cluster: 30	8
happy	3 (6.82)	happy	3 (6.82)	anger	2 (4.55)	happy	1 (2.33)	happy	2 (4.88)	Cluster: 31	25
mixed	1 (2.27)	anger	1 (2.27)	sad	1 (2.27)	mixed	1 (2.33)	mixed	2 (4.88)	Cluster: 32	25
						surprise	1 (2.33)	surprise	1 (2.33)	Cluster: 33	25
										Cluster: 34	25
										Cluster: 35	25
										Cluster: 36	25
										Cluster: 37	25
										Cluster: 38	25
										Cluster: 39	25
										Cluster: 40	25
										Cluster: 41	25
										Cluster: 42	25
										Cluster: 43	25
										Cluster: 44	25
										Cluster: 45	25
										Cluster: 46	25
										Cluster: 47	25
										Cluster: 48	25
										Cluster: 49	25
										Cluster: 50	25
										Cluster: 51	25
										Cluster: 52	25
										Cluster: 53	25
										Cluster: 54	25
										Cluster: 55	25
										Cluster: 56	25
										Cluster: 57	25
										Cluster: 58	25
										Cluster: 59	25
										Cluster: 60	25
										Cluster: 61	25
										Cluster: 62	25
										Cluster: 63	25
										Cluster: 64	25
										Cluster: 65	25
										Cluster: 66	25
										Cluster: 67	25
										Cluster: 68	25
										Cluster: 69	25
										Cluster: 70	25
										Cluster: 71	25
										Cluster: 72	25
										Cluster: 73	25
										Cluster: 74	25
										Cluster: 75	25
										Cluster: 76	25
										Cluster: 77	25
										Cluster: 78	25
										Cluster: 79	25
										Cluster: 80	25
										Cluster: 81	25
										Cluster: 82	25
										Cluster: 83	25
										Cluster: 84	25
										Cluster: 85	25
										Cluster: 86	25
										Cluster: 87	25
										Cluster: 88	25
										Cluster: 89	25
										Cluster: 90	25
										Cluster: 91	25
										Cluster: 92	25
										Cluster: 93	25
										Cluster: 94	25
										Cluster: 95	25
										Cluster: 96	25
										Cluster: 97	25
										Cluster: 98	25
										Cluster: 99	25
										Cluster: 100	25
										Cluster: 101	25
										Cluster: 102	25
										Cluster: 103	25
										Cluster: 104	25
										Cluster: 105	25
										Cluster: 106	25
										Cluster: 107	25
										Cluster: 108	25
										Cluster: 109	25
										Cluster: 110	25
										Cluster: 111	25
										Cluster: 112	25
										Cluster: 113	25
										Cluster: 114	25
										Cluster: 115	25
										Cluster: 116	25
										Cluster: 117	25
										Cluster: 118	25
										Cluster: 119	25
										Cluster: 120	25
										Cluster: 121	25
										Cluster: 122	25
										Cluster: 123	25
										Cluster: 124	25
										Cluster: 125	25
										Cluster: 126	25
										Cluster: 127	25
										Cluster: 128	25
										Cluster: 129	25
										Cluster: 130	25
										Cluster: 131	25
										Cluster: 132	25
										Cluster: 133	25
										Cluster: 134	25
										Cluster: 135	25
										Cluster: 136	25
										Cluster: 137	25
										Cluster: 138	25
										Cluster: 139	25
										Cluster: 140	25
										Cluster: 141	25
										Cluster: 142	25
										Cluster: 143	25
										Cluster: 144	25
										Cluster: 145	25
										Cluster: 146	25
										Cluster: 147	25
										Cluster: 148	25
										Cluster: 149	25
										Cluster: 150	25
										Cluster: 151	25
										Cluster: 152	25
										Cluster: 153	25
										Cluster: 154	25
										Cluster: 155	25
										Cluster: 156	25
										Cluster: 157	25
										Cluster: 158	25
										Cluster: 159	25
										Cluster: 160	25
										Cluster: 161	25

		Cluster Index: Total foci in that cluster Emotion Frequency of emotion (% of total cluster foci)							
Cluster: 31	7	Cluster: 32	7	Cluster: 33	5	Cluster: 34	5		
sad	3 (42.86)	disgust	3 (42.86)	aff	2 (40)	sad	2 (40)		
anger	2 (28.57)	aff	1 (14.29)	happy	1 (20)	disgust	1 (20)		
aff	1 (14.29)	anger	1 (14.29)	mixed	1 (20)	happy	1 (20)		
fear	1 (14.29)	fear	1 (14.29)	sad	1 (20)	mixed	1 (20)		
		happy	1 (14.29)						

* aff: affective.

TABLE 7

Meta-data cluster results using K-means.

Individual Foci Clusters				
Cluster Centers	Brain Regions	Cluster Index	# of foci per cluster (% of total foci)	# of studies per cluster (% of all studies)
(-27.28,-0.19,-19.05)	Amygdala L	1	192 (7.75)	93 (21.28)
(22.77,-5.71,-11.72)	Hippocampus R	2	187 (7.55)	81 (18.54)
(0.25,40.41,29.08)	Frontal Sup Medial L	3	169 (6.82)	82 (18.76)
(-1.63,-12.42,6.01)	Thalamus L	4	150 (6.05)	61 (13.96)
(-36.33,32.16,-3.46)	Frontal Inf Orb L	5	145 (5.85)	77 (17.62)
(5.35,35.91,-5.57)	Cingulum Ant R	6	134 (5.41)	60 (13.73)
(32.09,-56.43,-21.02)	Cerebelum 6 R	7	132 (5.33)	65 (14.87)
(42.2,13.91,-17.98)	Temporal Pole Sup R	8	131 (5.29)	65 (14.87)
(43.25,25.94,13.5)	Frontal Inf Tri R	9	131 (5.29)	61 (13.96)
(-2.38,-85.17,-3.49)	Calcarine L	10	123 (4.96)	54 (12.36)
(47.22,-61.42,7.4)	Temporal Mid R	11	121 (4.88)	56 (12.81)
(-36.22,-66,-19.17)	Cerebelum 6 L	12	108 (4.36)	52 (11.9)
(0.48,3.82,52.69)	Supp Motor Area L	13	105 (4.24)	57 (13.04)
(44.82,-11.83,22.29)	Rolandic Oper R	14	102 (4.12)	43 (9.84)
(4.57,-58.63,33.8)	Precuneus L	15	101 (4.08)	46 (10.53)
(-39.73,0.6,3.34)	Insula L	16	99 (4)	50 (11.44)
(-6.12,-40.69,-16.93)	Cerebelum 4 5 L	17	94 (3.79)	44 (10.07)
(-37.37,-67.84,22.79)	Temporal Mid L	18	86 (3.47)	45 (10.3)
(-49.48,-31.56,11.59)	Temporal Sup L	19	85 (3.43)	42 (9.61)
(-40.82,7.34,38.09)	Precentral L	20	83 (3.35)	45 (10.3)

R: right hemisphere, L: left hemisphere.