Modeling Inter-Subject Variability in Activation Locations of fMRI Data: A Bayesian Hierarchical Spatial Modeling Approach

Timothy D. Johnson[1]

Lei Xu[2]

Thomas E. Nichols[3]

Derek E. Nee[4]

University of Michigan[1]

Vanderbilt University[2]

GSK & University of Oxford[3]

Indiana University[4]

Outline

- Motivating example
- Statistical preliminaries
- Model Overview
- Model Details
- Results
- Conclusion

Motivating Example

Study of Proactive Interference Resolution

- Proactive interference is the phenomenon that recently learned information is mixed up with previously learned, similar, information
  - Information in this talk may get mixed up with similar information in the previous talk

Motivating Example

Study of Proactive Interference Resolution

- Proactive interference is the phenomenon that recently learned information is mixed up with previously learned, similar, information
  - Information in this talk may get mixed up with similar information in the previous talk
- One's ability to resolve proactive interference is key to in determining how much information one can store in short term memory

Motivating Example

Study of Proactive Interference Resolution

- Proactive interference is the phenomenon that recently learned information is mixed up with previously learned, similar, information
  - Information in this talk may get mixed up with similar information in the previous talk
- One’s ability to resolve proactive interference is key to in determining how much information one can store in short term memory
- 21 right-handed subjects participated in this study

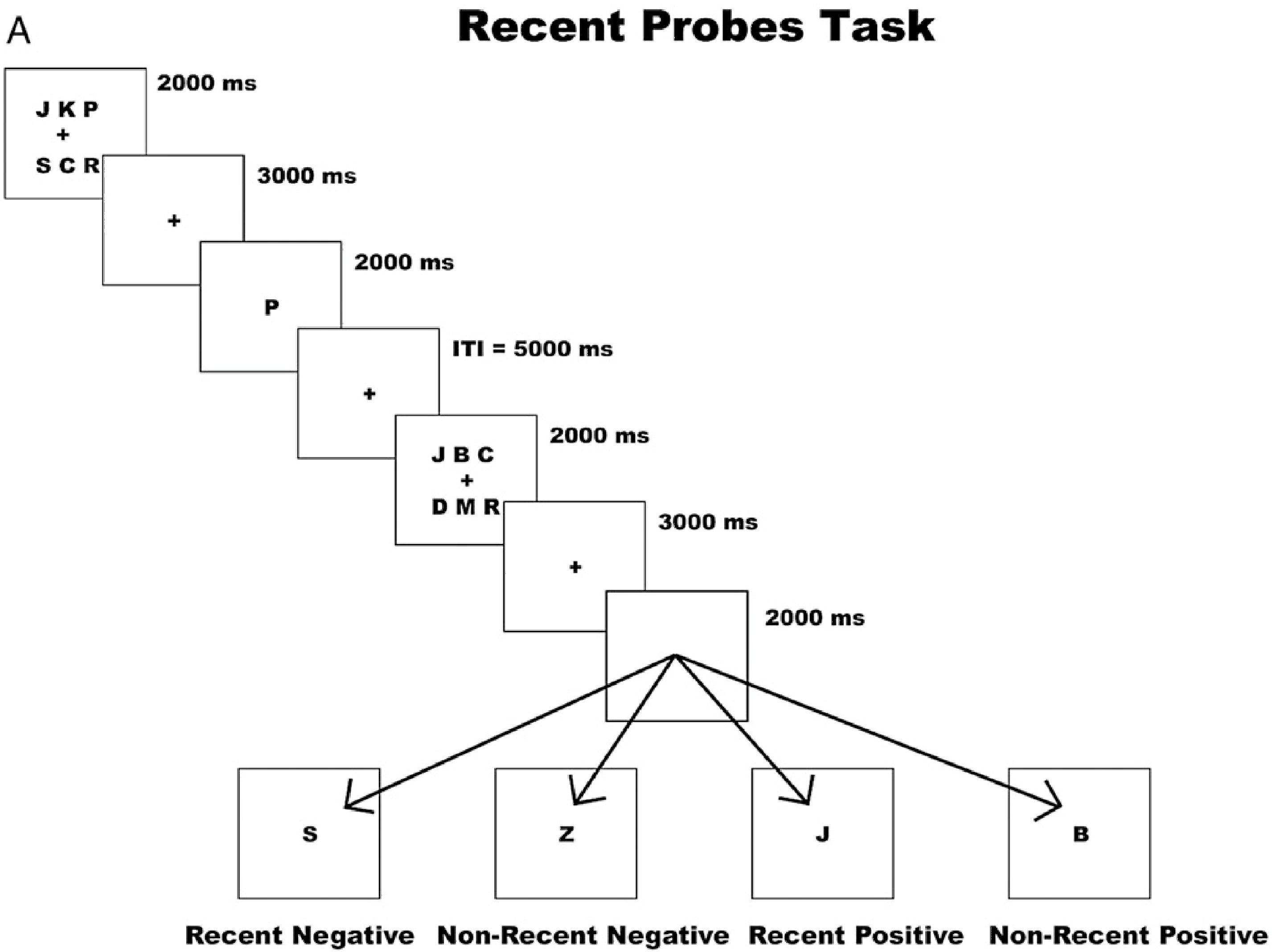
Motivating Example

Study of Proactive Interference Resolution

- Proactive interference is the phenomenon that recently learned information is mixed up with previously learned, similar, information
  - Information in this talk may get mixed up with similar information in the previous talk
- One’s ability to resolve proactive interference is key to in determining how much information one can store in short term memory
- 21 right-handed subjects participated in this study
  - We analyze the data from 18 subjects
  - 3 removed due to severe spiral artifacts

Nee, Jonides, Berman (2007), *Neuroimage*

Motivating Example



Motivating Example

- Recent probes task
  - Subjects show slower reaction time and increased error rates when rejecting *recent negative probes* compared to *non-recent negative probes*

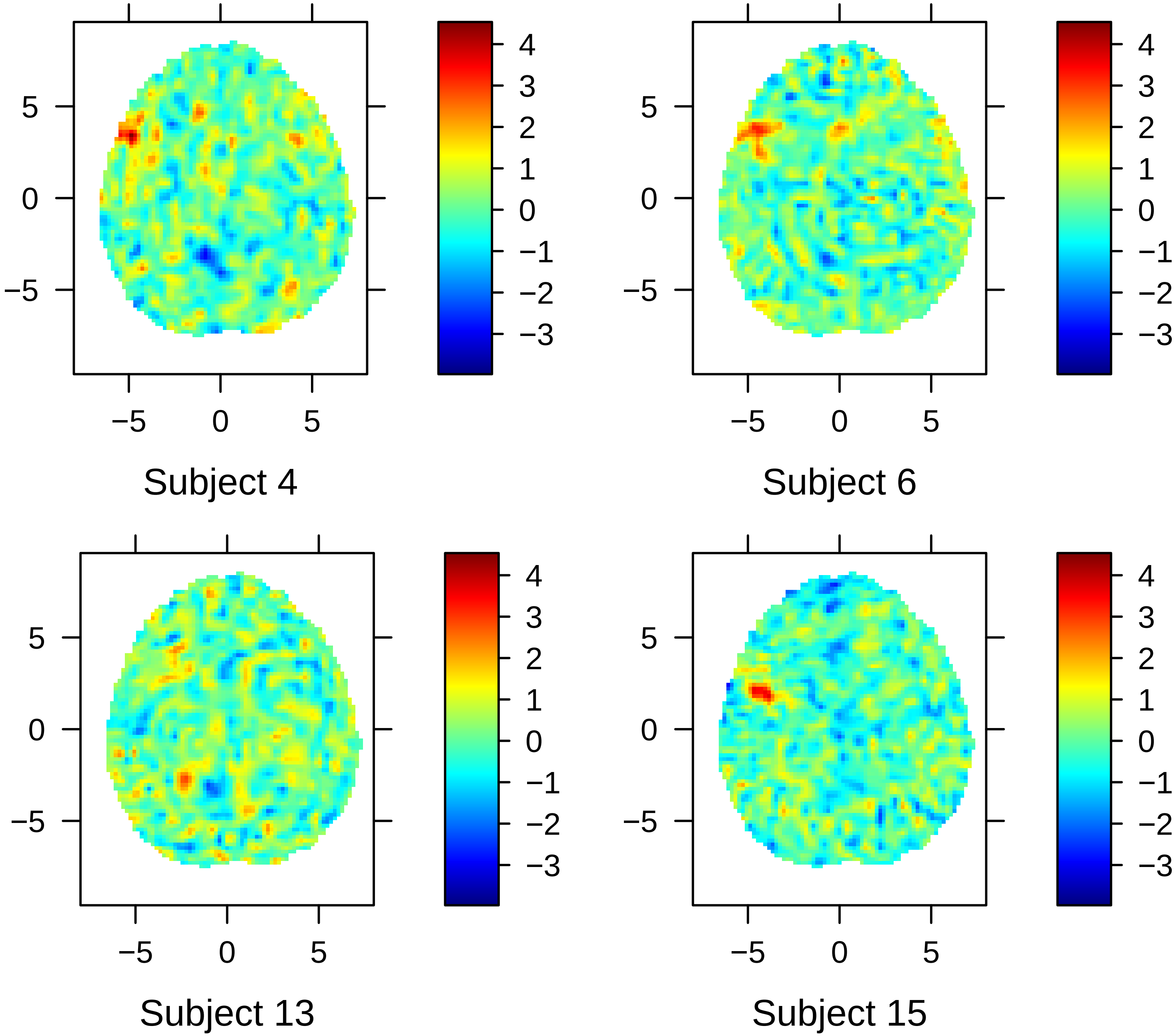
Motivating Example

- Recent probes task
  - Subjects show slower reaction time and increased error rates when rejecting *recent negative probes* compared to *non-recent negative probes*
  - Performance decrease a marker of proactive interference

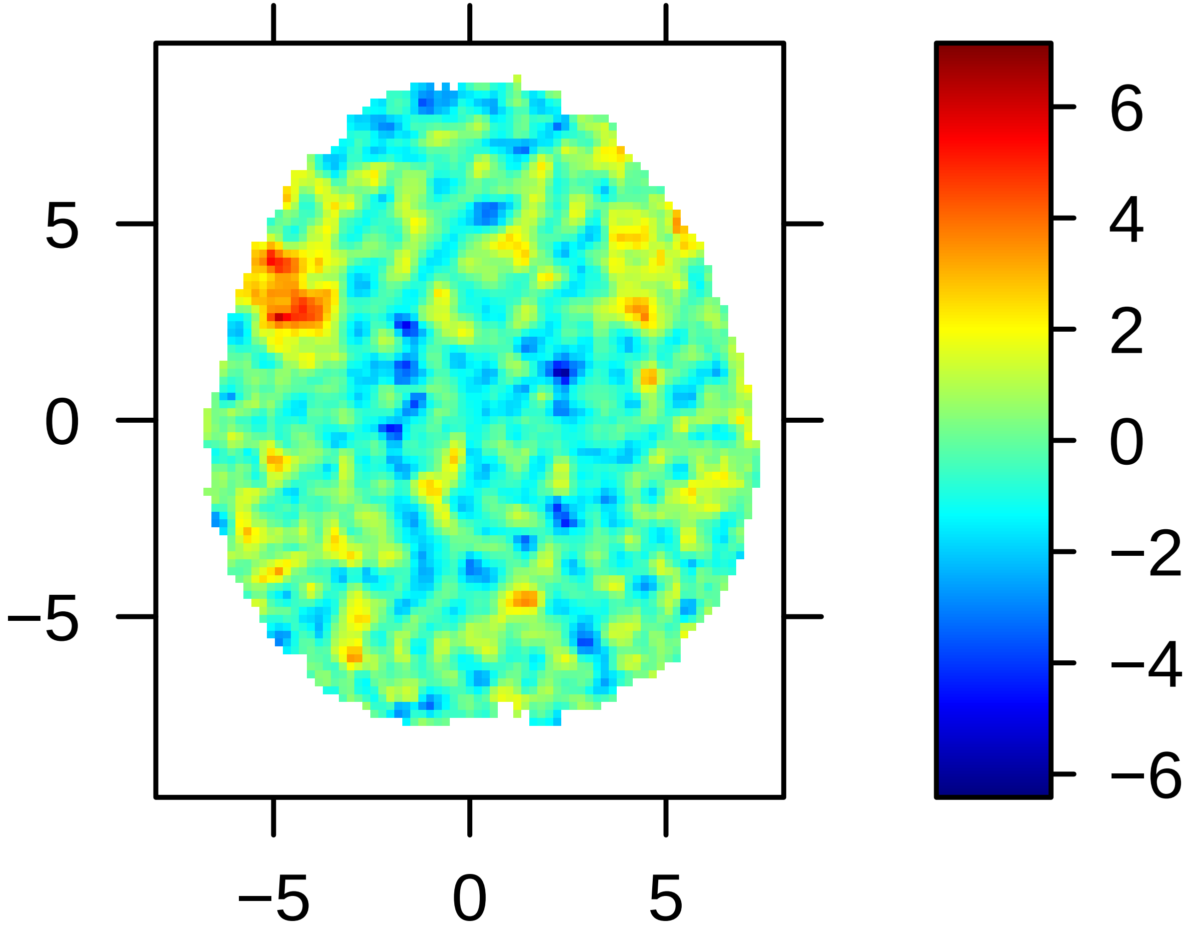
Motivating Example

- Recent probes task
  - Subjects show slower reaction time and increased error rates when rejecting *recent negative probes* compared to *non-recent negative probes*
  - Performance decrease a marker of proactive interference
- The left lateral prefrontal cortex is a region linked to proactive interference resolution

Motivating Example

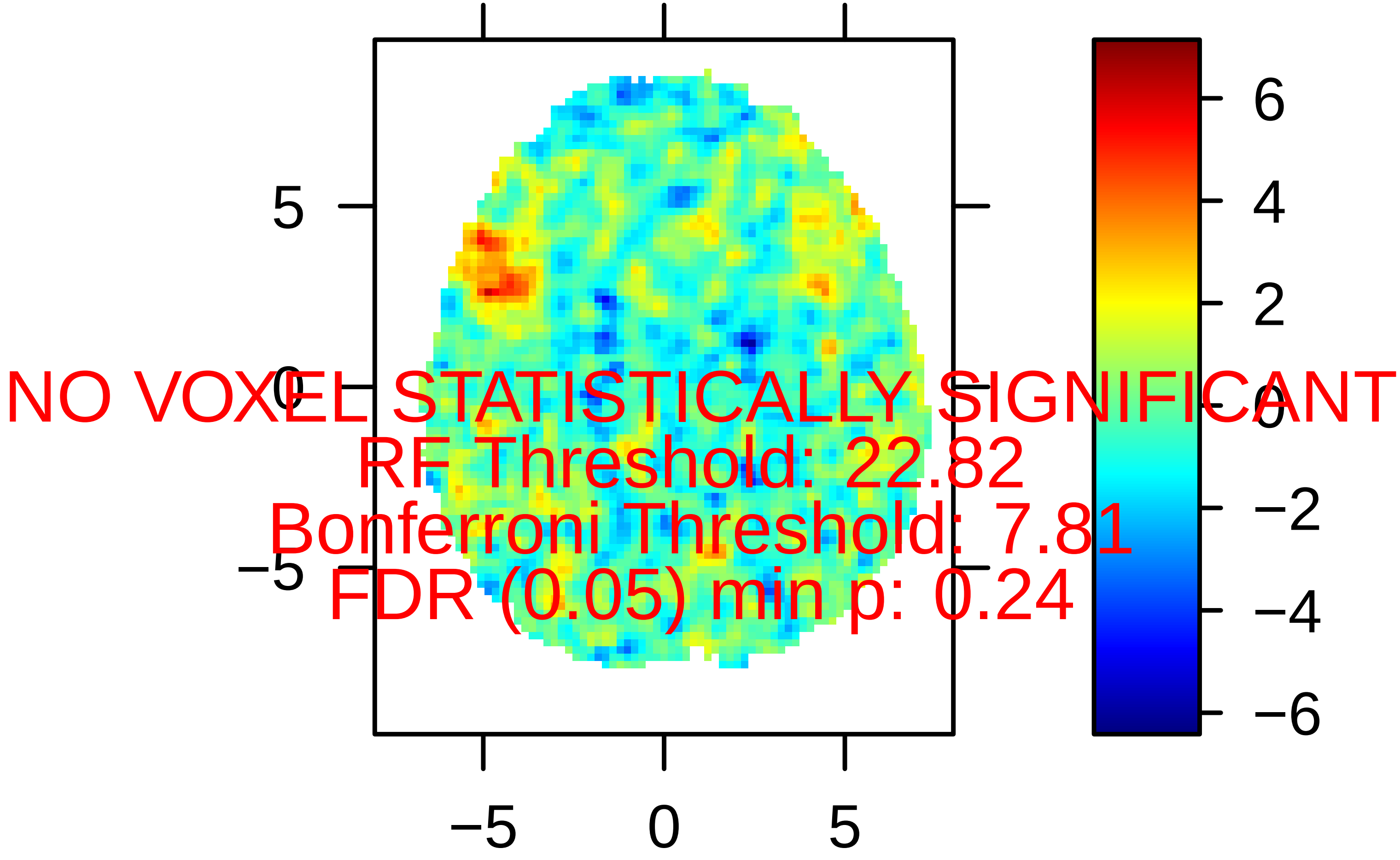


Motivating Example



Classical t−image

Motivating Example



Classical t−image

Preliminaries: Bayesian Statistics

- Central difference between *frequentist* and *Bayesian* paradigms
  - Bayesian paradigm: parameters considered random
  - Frequenist paradigm: parameters considered unknown constants
- All parameters equipped with a distribution (the prior distribution)
- Estimate the posterior distribution of the parameters given the data
  - The probability inversion is performed via Bayes Theorem:

$$\pi(\boldsymbol{\theta} \mid \mathbf{Y}) = \frac{\pi(\mathbf{Y} \mid \boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\pi(\mathbf{Y})} = \frac{\pi(\mathbf{Y}, \boldsymbol{\theta})}{\int \pi(\mathbf{Y}, \boldsymbol{\theta})d\boldsymbol{\theta}}$$

Preliminaries: Finite Mixture Distributions

- Suppose a population is made up of several sub-populations
- $Y \sim F_i(\theta_i)$ for sub-population i, $i = 1, \dots, n$
- Suppose that sub-population i makes up p_i of the total population with $\sum_i p_i = 1$
- Then the population has the mixture distribution:

$$Y \sim \sum_{i=1}^{n} p_i \, F_i(\theta_i)$$

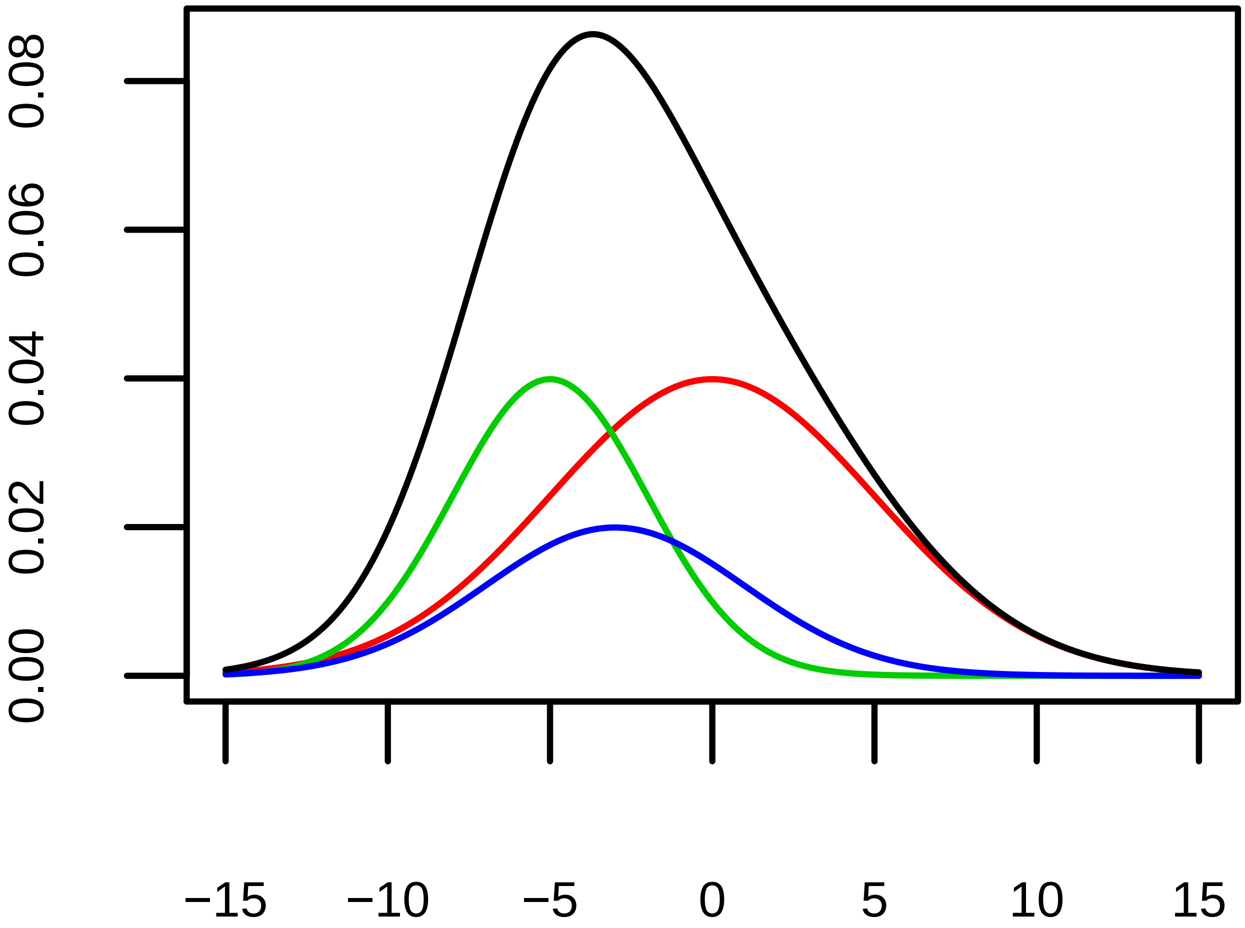
with components $F_i(\theta_i)$ and weights p_i

- Define a latent allocation variable Z_j with $\Pr(Z_j = k) = p_k$

$$[Y_j \mid Z_j = k] \sim F_k(\theta_k)$$

- RJMCMC

Preliminaries: Finite Mixture Distributions



Preliminaries: Cox Cluster Process

- Given a set of points $\{y_j\}_{j=1}^m$ in $\mathbb{R}^3$, we assume that they are a realization of a spatial Poisson process with intensity

$$\lambda(y \mid \{x_i\}_{i=1}^n) = \epsilon + \sum_{i=1}^n h(y \mid x_i)$$

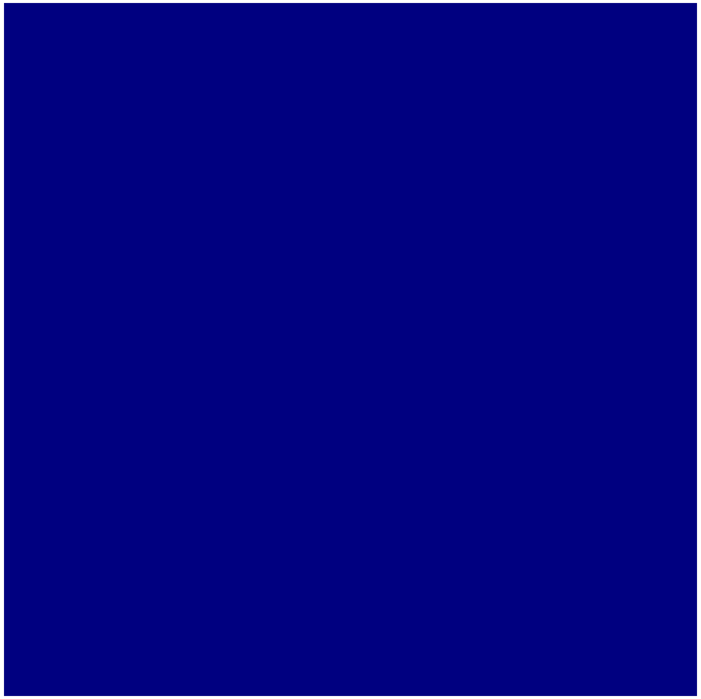
- ϵ is the underlying background intensity
- $h(y \mid x_i) : \mathbb{R}^3 \to \mathbb{R}^+ \cup \{0\}$ is some non-negative function
- We place, a priori, a (marked) Poisson process on $\{x_i\}$.
- $\{x_i\}$ can also be repulsive. e.g. a hard core process:

$$\pi(\{x_i\}) = \begin{cases} \alpha\beta^{|\{x_i\}|} & \text{if} \quad ||x_i - x_j|| > R \; \forall \; i \neq j \\ 0 & \end{cases}$$

for some fixed radius $R > 0$ and intensity $\beta > 0$

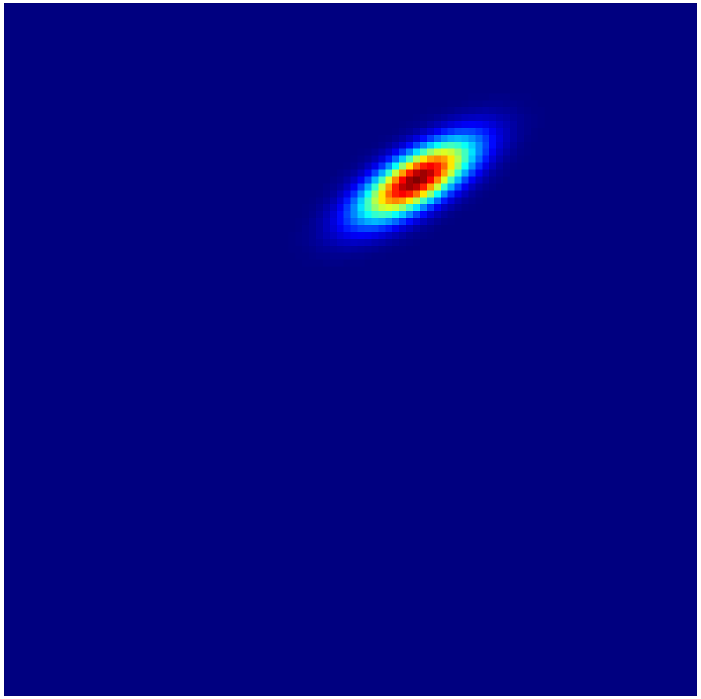
Preliminaries: Cox Cluster Process

$$\lambda(y \mid \emptyset) = \epsilon$$



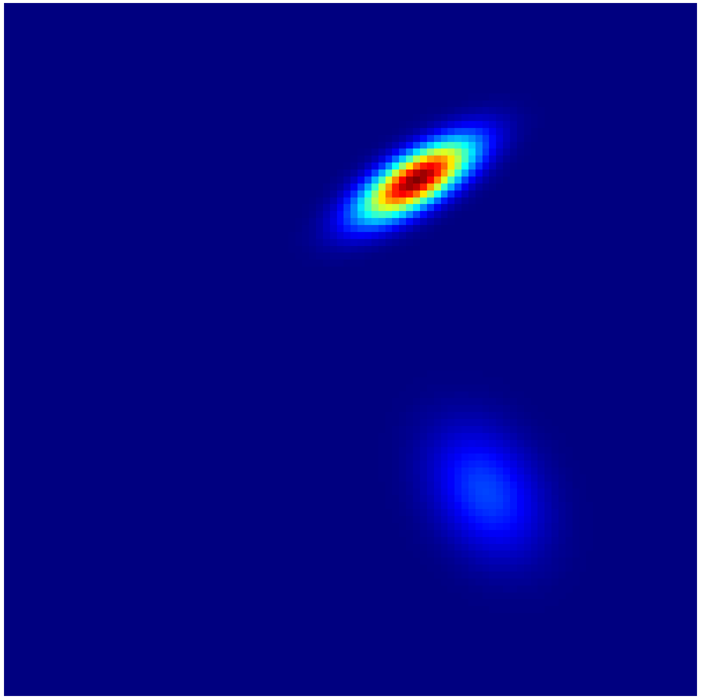
Preliminaries: Cox Cluster Process

$$\lambda(y \mid \{x_1\}) = \epsilon + h(y \mid x_1)$$



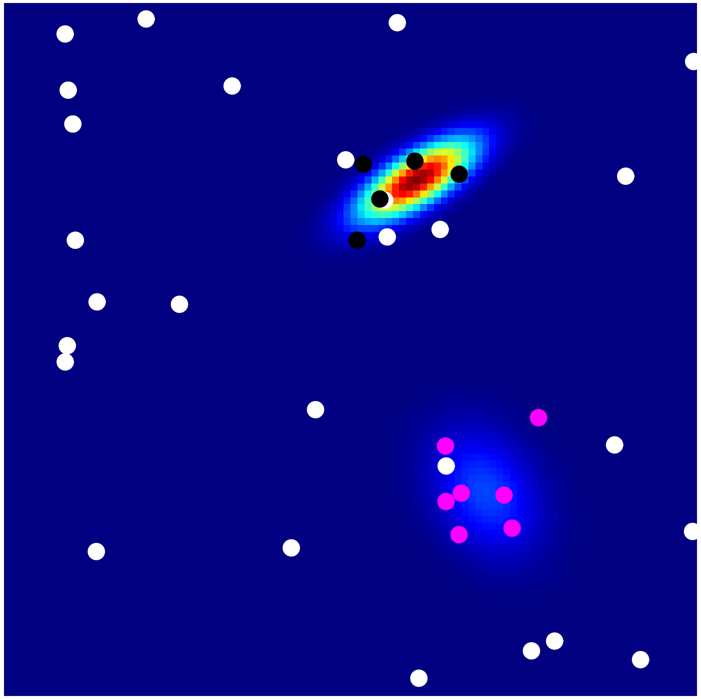
Preliminaries: Cox Cluster Process

$$\lambda(y \mid \{x_1, x_2\}) = \epsilon + h(y \mid x_1) + h(y \mid x_2)$$



Preliminaries: Cox Cluster Process

A particular instance of this process may look like...



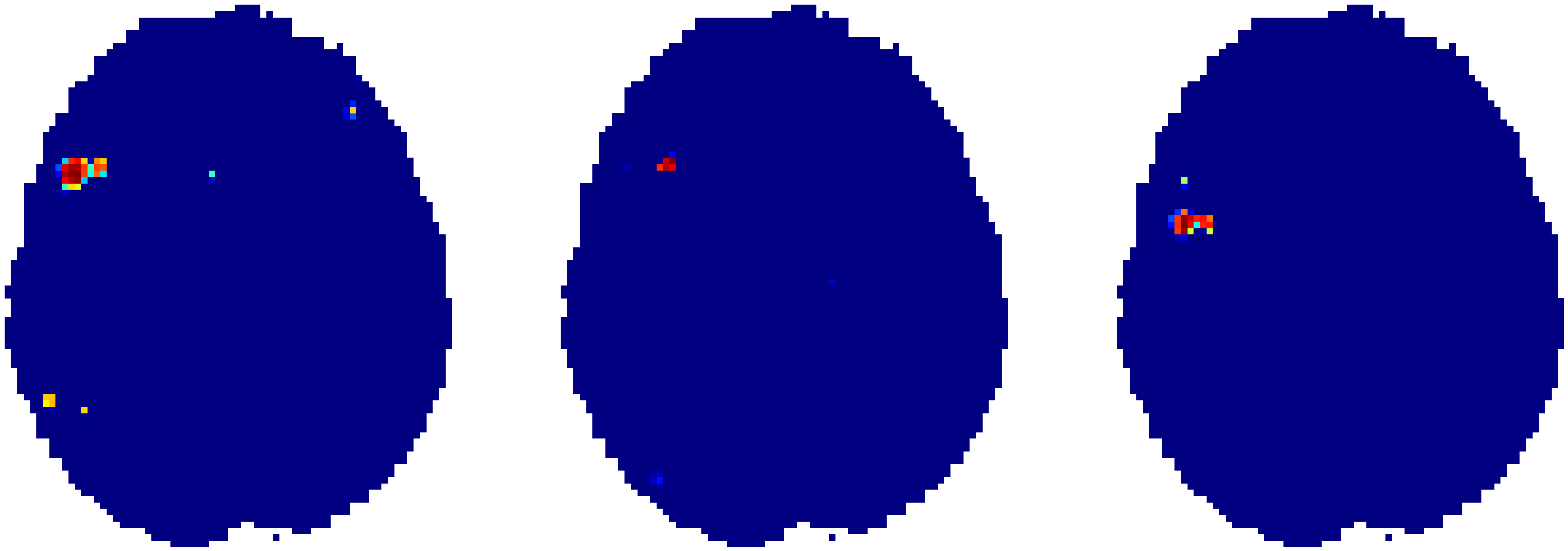
Model Overview

A Bayesian Spatial Hierarchical Model

Model Overview

A Bayesian Spatial Hierarchical Model

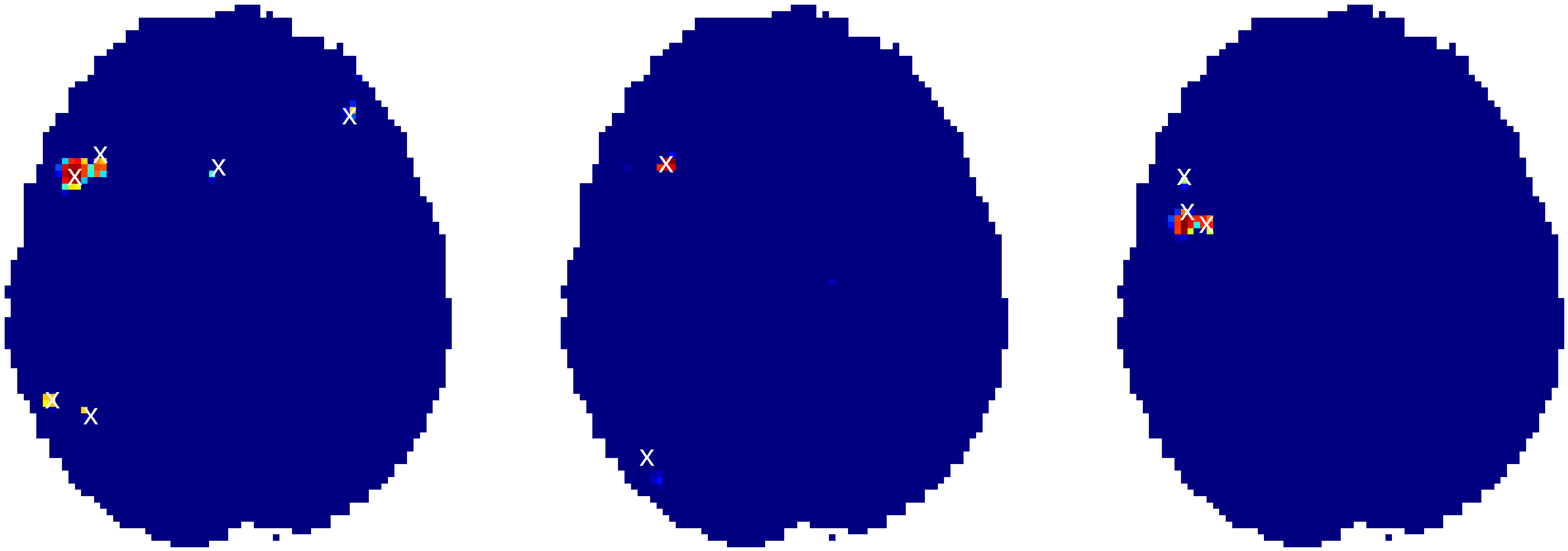
- Level 1: subject level data
  - BOLD image modeled as a mixture distribution
  - Spatial correlation accounted for in the mixing weights



Model Overview

A Bayesian Spatial Hierarchical Model

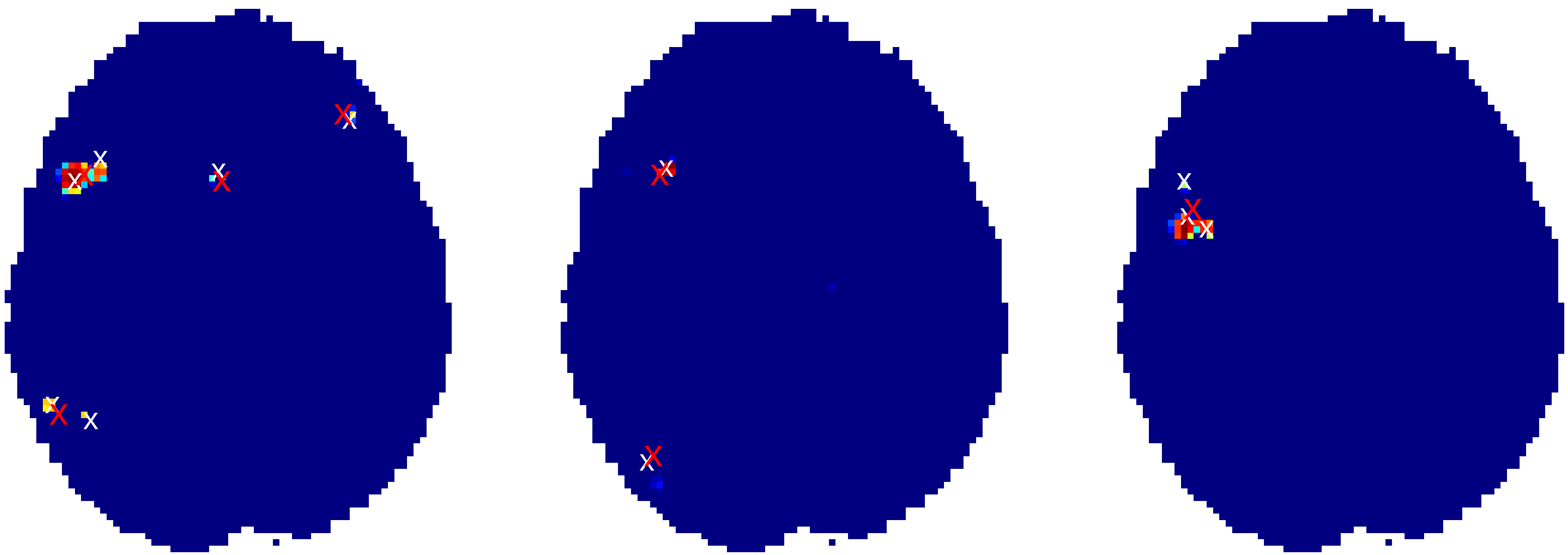
- Level 1: subject level data
  - BOLD image modeled as a mixture distribution
  - Spatial correlation accounted for in the mixing weights



Model Overview

A Bayesian Spatial Hierarchical Model

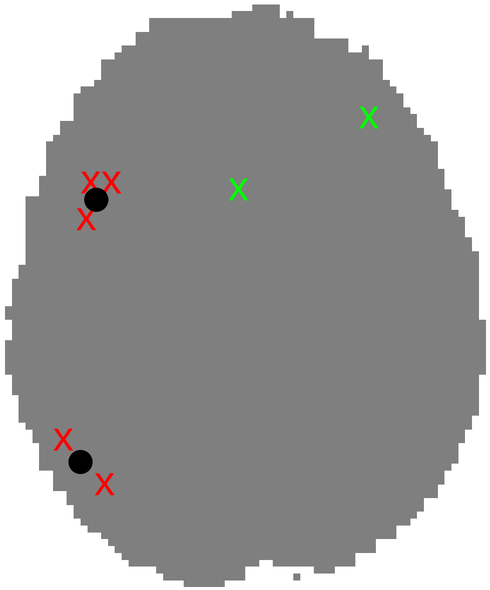
- Level 2: subject level data
  - Component means clustered about “activation centers”
  - Again, a mixture distribution is used



Model Overview

A Bayesian Spatial Hierarchical Model

- Level 3: population level data
  - Activation centers clustered around pop level centers
  - Via a spatial Cox cluster process



Model Overview

A Bayesian Spatial Hierarchical Model

- Level 4: population level data
  - Population centers equipped with a homogeneous spatial Poisson process
  - Therefore, conditional on the number of centers
    - The population centers are, a priori, iid uniformly throughout the brain

Model Details

- Level 1:
  - Data assumed to come from a mixture distribution

$$[Y_{ij} \mid \cdot] \sim p_{ij0} N(\theta_0, \sigma_0^2) + \sum_{\ell=1}^{c_i} p_{ij\ell} N(\theta_{i\ell}, \sigma_{i\ell}^2)$$

Model Details

- Level 1:
  - Data assumed to come from a mixture distribution

$$[Y_{ij} \mid \cdot] \sim p_{ij0} N(\theta_0, \sigma_0^2) + \sum_{\ell=1}^{c_i} p_{ij\ell} N(\theta_{i\ell}, \sigma_{i\ell}^2)$$

  - Spatial correlation accounted for in the mixing weights

$$p_{ij\ell} \propto \begin{cases} q & \ell = 0 \\ \phi_3(\mathbf{x}_{ij}; \boldsymbol{\eta}_{i\ell}, \Psi_{i\ell}) & \ell = 1, \ldots, c_i \end{cases}$$

where $\sum_{\ell=0}^{c_i} p_{ij\ell} = 1$

Model Details

- Level 1:
  - Data assumed to come from a mixture distribution

$$[Y_{ij} \mid \cdot] \sim p_{ij0} N(\theta_0, \sigma_0^2) + \sum_{\ell=1}^{c_i} p_{ij\ell} N(\theta_{i\ell}, \sigma_{i\ell}^2)$$

  - Spatial correlation accounted for in the mixing weights

$$p_{ij\ell} \propto \begin{cases} q & \ell = 0 \\ \phi_3(\mathbf{x}_{ij}; \boldsymbol{\eta}_{i\ell}, \Psi_{i\ell}) & \ell = 1, \ldots, c_i \end{cases}$$

where $\sum_{\ell=0}^{c_i} p_{ij\ell} = 1$

  - $\boldsymbol{\eta}_{i\ell}$ is the location mean of component ℓ, subject i

Model Details

- Level 1:
  - Data assumed to come from a mixture distribution

$$[Y_{ij} \mid \cdot] \sim p_{ij0} N(\theta_0, \sigma_0^2) + \sum_{\ell=1}^{c_i} p_{ij\ell} N(\theta_{i\ell}, \sigma_{i\ell}^2)$$

  - Spatial correlation accounted for in the mixing weights

$$p_{ij\ell} \propto \begin{cases} q & \ell = 0 \\ \phi_3(\mathbf{x}_{ij}; \boldsymbol{\eta}_{i\ell}, \Psi_{i\ell}) & \ell = 1, \ldots, c_i \end{cases}$$

where $\sum_{\ell=0}^{c_i} p_{ij\ell} = 1$

  - $\boldsymbol{\eta}_{i\ell}$ is the location mean of component ℓ, subject i
  - Define latent allocation variables w_{ij}: $\Pr(w_{ij} = \ell) = p_{ij\ell}$

Model Details

- Level 1 Priors:
  - $\theta_0 \sim N(0,1) \qquad \sigma_0^{-2} \sim G(0.001, 0.001)$

Model Details

- Level 1 Priors:
  - $\theta_0 \sim N(0,1) \qquad \sigma_0^{-2} \sim G(0.001, 0.001)$
  - $\theta_{i\ell} \sim N(3.92, 1) \qquad \sigma_{i\ell}^{-2} \sim G(3, \beta) \qquad \beta \sim G(0.001, 0.001)$
    - (densities intersect at 1.96, the 0.975 quantile of the standard normal)

Model Details

- Level 1 Priors:
  - $\theta_0 \sim N(0,1) \qquad \sigma_0^{-2} \sim G(0.001, 0.001)$
  - $\theta_{i\ell} \sim N(3.92, 1) \qquad \sigma_{i\ell}^{-2} \sim G(3, \beta) \qquad \beta \sim G(0.001, 0.001)$
    - (densities intersect at 1.96, the 0.975 quantile of the standard normal)
  - Experiments designed so only a small portion of the brain "activates"
    - Set $q = 39$ — controls number of activated sites

Model Details

- Level 1 Priors:
  - $\theta_0 \sim N(0,1) \qquad \sigma_0^{-2} \sim G(0.001, 0.001)$
  - $\theta_{i\ell} \sim N(3.92,1) \qquad \sigma_{i\ell}^{-2} \sim G(3,\beta) \qquad \beta \sim G(0.001, 0.001)$
    - (densities intersect at 1.96, the 0.975 quantile of the standard normal)
  - Experiments designed so only a small portion of the brain "activates"
    - Set $q = 39$ — controls number of activated sites
    - $\Psi_{i\ell}^{-1} \sim W(10, 0.2\pi I_{3\times 3})$
      - Results in a priori prob. of 0.975 that site $\mathbf{x}_{ij}$ belongs to background when $\mathbf{x}_{ij} \equiv \boldsymbol{\eta}_{i\ell}$ and all other η_{ik} "far" away

Model Details

- Level 1 Priors:
  - $\theta_0 \sim N(0,1) \qquad \sigma_0^{-2} \sim G(0.001, 0.001)$
  - $\theta_{i\ell} \sim N(3.92, 1) \qquad \sigma_{i\ell}^{-2} \sim G(3, \beta) \qquad \beta \sim G(0.001, 0.001)$
    - (densities intersect at 1.96, the 0.975 quantile of the standard normal)
  - Experiments designed so only a small portion of the brain "activates"
    - Set $q = 39$ — controls number of activated sites
    - $\Psi_{i\ell}^{-1} \sim W(10, 0.2\pi I_{3\times 3})$
      - Results in a priori prob. of 0.975 that site $\mathbf{x}_{ij}$ belongs to background when $\mathbf{x}_{ij} \equiv \boldsymbol{\eta}_{i\ell}$ and all other η_{ik} "far" away
  - $\Pr(c_i = K) = 1/200, \quad K = 1, \ldots, 200$
    - RJMCMC used to estimate the number of mixture components

(Green, P. (1995) Biometrika)

Model Details

- Level 2:
  - Component means distributed about activation “centers”
    - (may take several components to adequately fit large activation clusters)

$$\pi(\boldsymbol{\eta}_{i\ell} \mid \cdot) = \sum_{k=1}^{b_i} q_{ik} \frac{\phi_3(\boldsymbol{\eta}_{i\ell}; \boldsymbol{\nu}_{ik}, \Phi_{ik})}{\Pr(\boldsymbol{\eta}_{i\ell} \in B_i \mid \boldsymbol{\nu}_{ik}, \Phi_{ik})} 1_{B_i}(\boldsymbol{\eta}_{i\ell})$$

Model Details

- Level 2:
  - Component means distributed about activation "centers"
    - (may take several components to adequately fit large activation clusters)

$$\pi(\boldsymbol{\eta}_{i\ell} \mid \cdot) = \sum_{k=1}^{b_i} q_{ik} \frac{\phi_3(\boldsymbol{\eta}_{i\ell}; \boldsymbol{\nu}_{ik}, \Phi_{ik})}{\Pr(\boldsymbol{\eta}_{i\ell} \in B_i \mid \boldsymbol{\nu}_{ik}, \Phi_{ik})} 1_{B_i}(\boldsymbol{\eta}_{i\ell})$$

  - Latent allocation variables $v_{i\ell}$: $\Pr(v_{i\ell} = k) = q_{ik}$

Model Details

- Level 2:
  - Component means distributed about activation “centers”
    - (may take several components to adequately fit large activation clusters)

$$\pi(\boldsymbol{\eta}_{i\ell} \mid \cdot) = \sum_{k=1}^{b_i} q_{ik} \frac{\phi_3(\boldsymbol{\eta}_{i\ell}; \boldsymbol{\nu}_{ik}, \Phi_{ik})}{\Pr(\boldsymbol{\eta}_{i\ell} \in B_i \mid \boldsymbol{\nu}_{ik}, \Phi_{ik})} 1_{B_i}(\boldsymbol{\eta}_{i\ell})$$

  - Latent allocation variables $v_{i\ell}$: $\Pr(v_{i\ell} = k) = q_{ik}$
    - Note: component means $\boldsymbol{\eta}_{i\ell}$ and activation centers $\boldsymbol{\nu}_{ik}$ are latent as well, i.e. not observable

Model Details

- Level 2 Priors:
  - $[\mathbf{q}_i^{\mathrm{T}} = (q_{i1}, \ldots, q_{ib_i}) \mid b_i] \sim D(1, \ldots, 1)$

Model Details

- Level 2 Priors:
  - $[\mathbf{q}_i^{\mathrm{T}} = (q_{i1}, \ldots, q_{ib_i}) \mid b_i] \sim D(1, \ldots, 1)$
  - $\Phi_{ik}^{-1} \sim W(5, R^{-1}/3), \quad R \sim W(5, S/5), \quad S = 4\, I_{3\times 3}$
    - Results in $E(\Phi_{ik}) = S$ (FWHM ≈ 0.94 cm)
    - A priori, a 95% credible sphere of radius ≈ 1.0 cm

Model Details

- Level 2 Priors:
  - $[\mathbf{q}_i^{\mathrm{T}} = (q_{i1}, \ldots, q_{ib_i}) \mid b_i] \sim D(1, \ldots, 1)$
  - $\Phi_{ik}^{-1} \sim W(5, R^{-1}/3), \quad R \sim W(5, S/5), \quad S = 4\, I_{3\times 3}$
    - Results in $E(\Phi_{ik}) = S$ (FWHM ≈ 0.94 cm)
    - A priori, a 95% credible sphere of radius ≈ 1.0 cm
  - $\Pr(b_i = M) = 1/50, \quad M = 1, \ldots, 50$
    - RJMCMC used to estimate # of activation centers for each individual

Model Details

- Level 2 Priors:
  - $[\mathbf{q}_i^{\mathrm{T}} = (q_{i1}, \ldots, q_{ib_i}) \mid b_i] \sim D(1, \ldots, 1)$
  - $\Phi_{ik}^{-1} \sim W(5, R^{-1}/3), \quad R \sim W(5, S/5), \quad S = 4\, I_{3\times 3}$
    - Results in $E(\Phi_{ik}) = S$ (FWHM ≈ 0.94 cm)
    - A priori, a 95% credible sphere of radius ≈ 1.0 cm
  - $\Pr(b_i = M) = 1/50, \quad M = 1, \ldots, 50$
    - RJMCMC used to estimate # of activation centers for each individual
- So far, all modeling done at the subject level

Model Details

- Level 3:
    - Individual activation centers are clustered about population activation centers via a spatial Cox cluster process

Model Details

- Level 3:
  - Individual activation centers are clustered about population activation centers via a spatial Cox cluster process
  - Intensity function given by

$$\lambda\left(\boldsymbol{\nu}_{ik} \mid \{(\boldsymbol{\mu}_i, \Sigma_i)\}_{i=1}^{N}\right) = \epsilon + \theta \sum_{i=1}^{N} \frac{\phi_3(\boldsymbol{\nu}_{ik}; \boldsymbol{\mu}_i, \Sigma_i)}{\Pr(\boldsymbol{\nu}_{ik} \in B_i; \boldsymbol{\mu}_i, \Sigma_i)} 1_{B_i}(\boldsymbol{\nu}_{ik})$$

Model Details

- Level 3:
  - Individual activation centers are clustered about population activation centers via a spatial Cox cluster process
  - Intensity function given by

$$\lambda\left(\boldsymbol{\nu}_{ik} \mid \{(\boldsymbol{\mu}_i, \Sigma_i)\}_{i=1}^{N}\right) = \epsilon + \theta \sum_{i=1}^{N} \frac{\phi_3(\boldsymbol{\nu}_{ik}; \boldsymbol{\mu}_i, \Sigma_i)}{\Pr(\boldsymbol{\nu}_{ik} \in B_i; \boldsymbol{\mu}_i, \Sigma_i)} 1_{B_i}(\boldsymbol{\nu}_{ik})$$

  - Conditional on the number, N_A, of individual activation centers, their locations, $\boldsymbol{\nu}_{ik}$, are iid with uniform distribution over the volume of the brain

$$[\boldsymbol{\nu}_{ik} \mid N_a] \sim U[V(B_i)]$$

Model Details

- Level 3 Priors:
  - $\epsilon \sim G(54, V(B)) \Rightarrow E(N_s) = 54$ where $B = \cup B_i$
    - N_s denotes the # of spurious ind. act. centers, works out to an expected number of 3 spurious activation centers per subject
    - Placing a hyperprior dist. on ϵ reflects our uncertainty in the expected number

Model Details

- Level 3 Priors:
  - $\epsilon \sim G(54, V(B)) \Rightarrow E(N_s) = 54$ where $B = \cup B_i$
    - N_s denotes the # of spurious ind. act. centers, works out to an expected number of 3 spurious activation centers per subject
    - Placing a hyperprior dist. on ϵ reflects our uncertainty in the expected number
  - $\theta \sim G(9, 1)$
    - We except, a priori, on average about half the subjects will have an activation center cluster about any given population center

Model Details

- Level 4:
  - $\{(\boldsymbol{\mu}_i, \Sigma_i)\}$ a marked homogeneous Poisson process with intensity λ
    - i.e. $\{\boldsymbol{\mu}_i\}$ is a homo. Poisson process with marks $\{\Sigma_i\}$

Model Details

- Level 4:
  - $\{(\boldsymbol{\mu}_i, \Sigma_i)\}$ a marked homogeneous Poisson process with intensity λ
    - i.e. $\{\boldsymbol{\mu}_i\}$ is a homo. Poisson process with marks $\{\Sigma_i\}$
- Level 4 Priors:
  - $\lambda = 10$

Model Details

- Level 4:
  - $\{(\boldsymbol{\mu}_i, \Sigma_i)\}$ a marked homogeneous Poisson process with intensity λ
    - i.e. $\{\boldsymbol{\mu}_i\}$ is a homo. Poisson process with marks $\{\Sigma_i\}$
- Level 4 Priors:
  - $\lambda = 10$
  - $\Sigma_i^{-1} \sim W(5, T^{-1}/3)$
    - $T \sim W(5, D/3)$
    - $D = 6.25 I_{3\times 3}$
      - Results in $E(\Sigma_i) = D$, a priori (FWHM ≈ 1.18 cm)
      - A 95% credible sphere of radius ≈ 1.4 cm

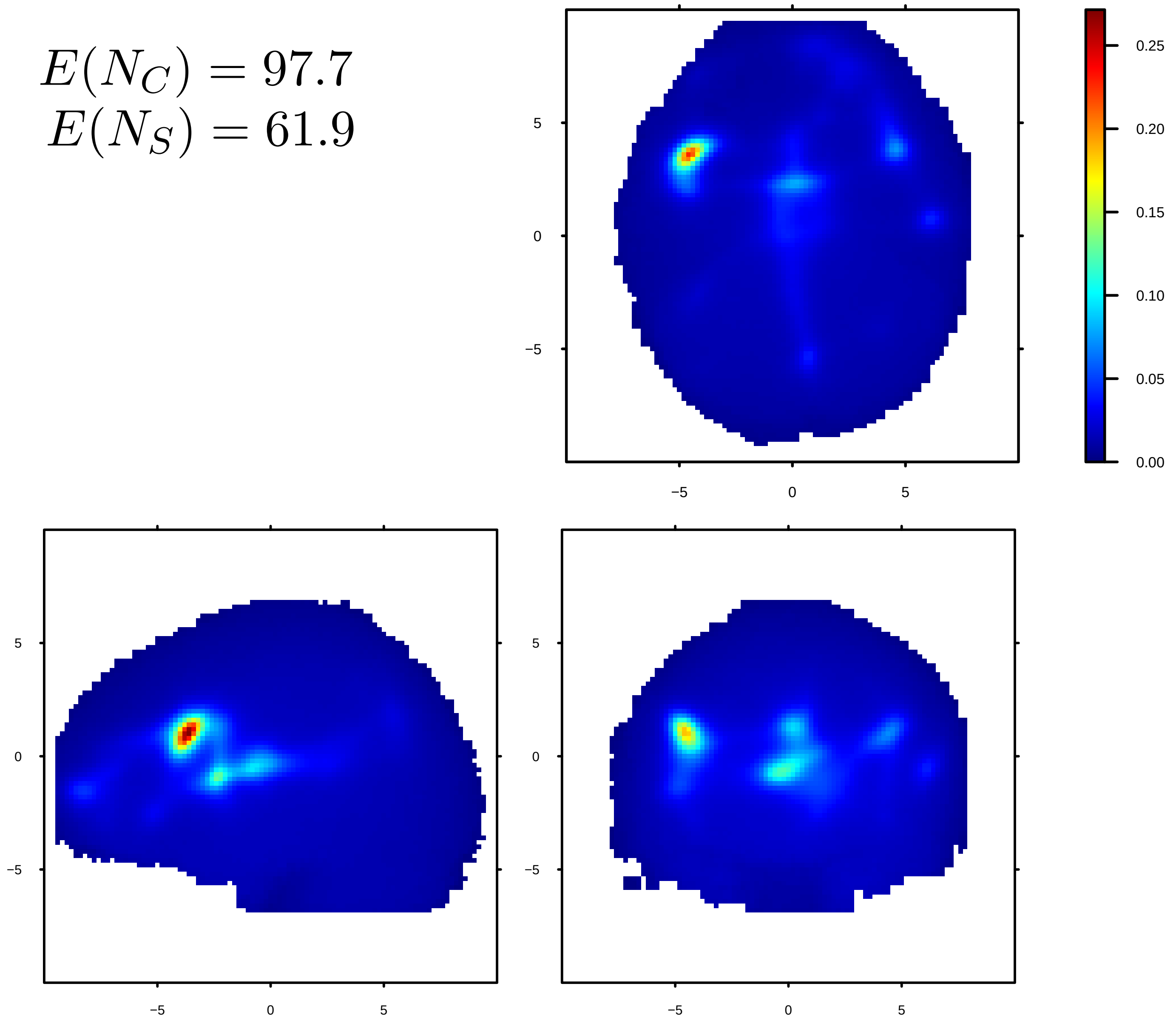
Model Details

- Level 4:
  - $\{(\boldsymbol{\mu}_i, \Sigma_i)\}$ a marked homogeneous Poisson process with intensity λ
    - i.e. $\{\boldsymbol{\mu}_i\}$ is a homo. Poisson process with marks $\{\Sigma_i\}$
- Level 4 Priors:
  - $\lambda = 10$
  - $\Sigma_i^{-1} \sim W(5, T^{-1}/3)$
    - $T \sim W(5, D/3)$
    - $D = 6.25 I_{3\times 3}$
      - Results in $E(\Sigma_i) = D$, a priori (FWHM ≈ 1.18 cm)
      - A 95% credible sphere of radius ≈ 1.4 cm
- Posterior of $\{\mu_i\}$ simulated via a spatial birth-death process

van Lieshout & Baddeley (2002), in *Spatial Cluster Modelling*, Ch 4.

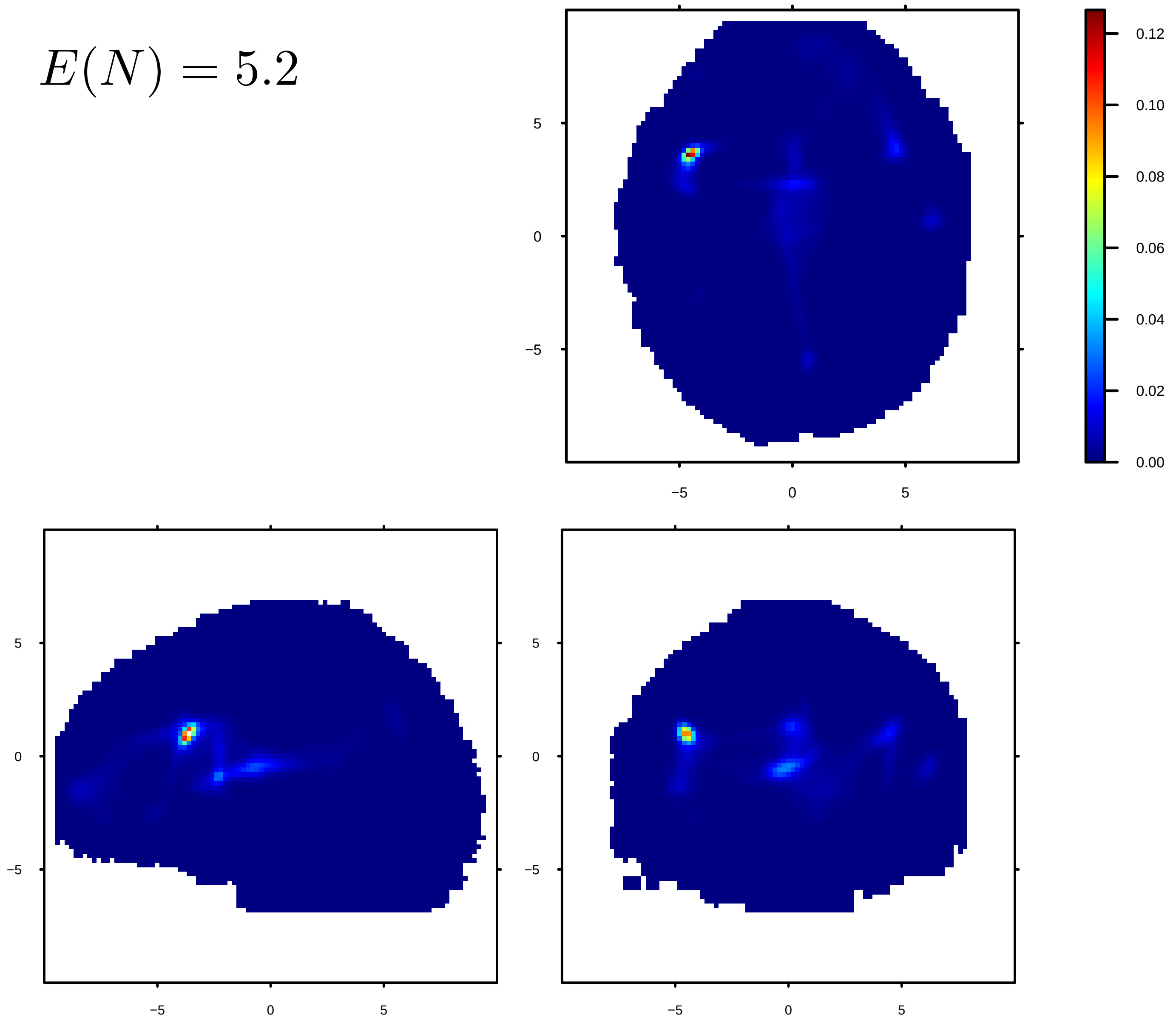
Results: Marginal intensity of Ind Centers

$$E(N_C) = 97.7$$
$$E(N_S) = 61.9$$



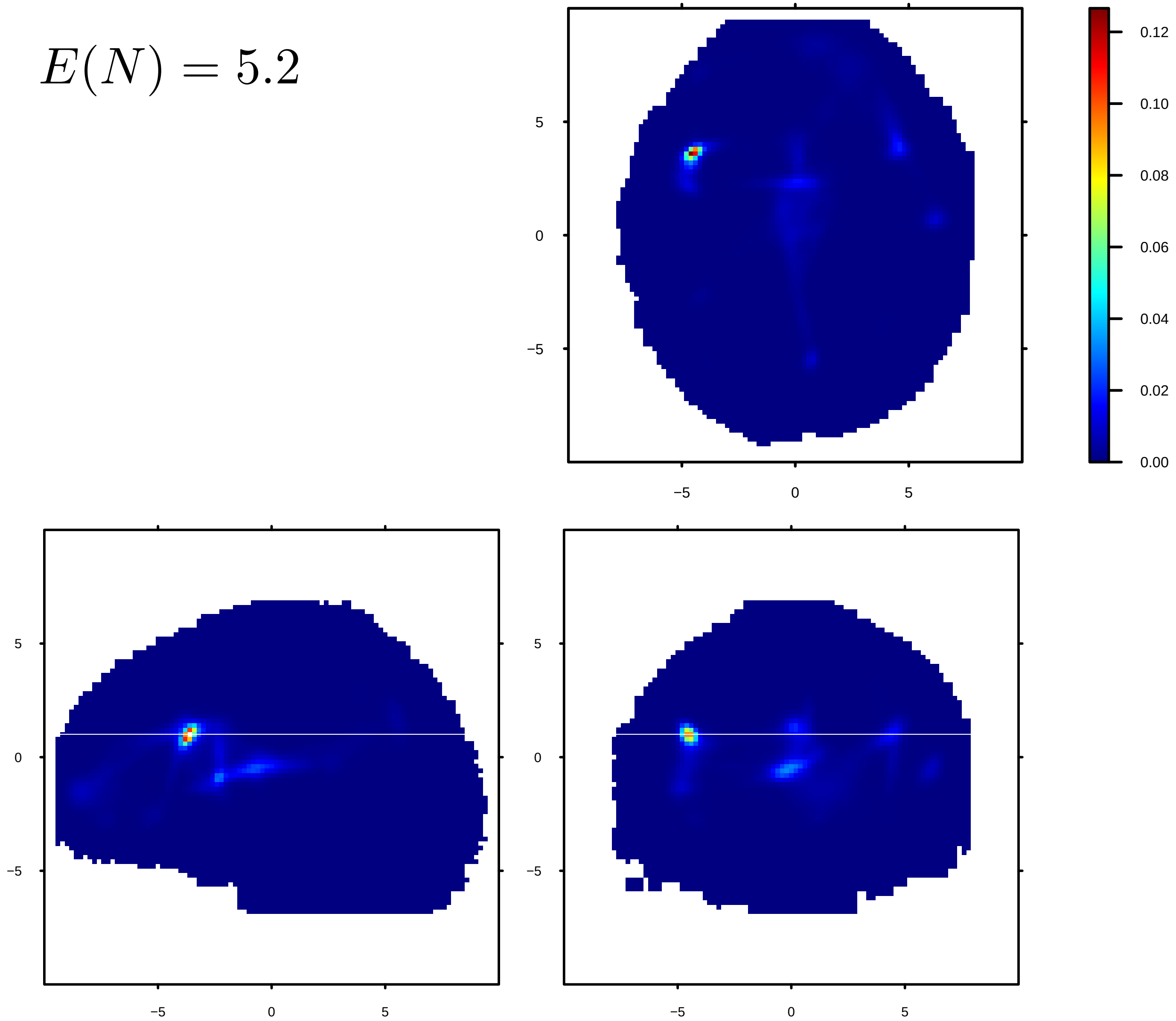
Results: Marginal intensity of Pop Ctrs

$$E(N) = 5.2$$

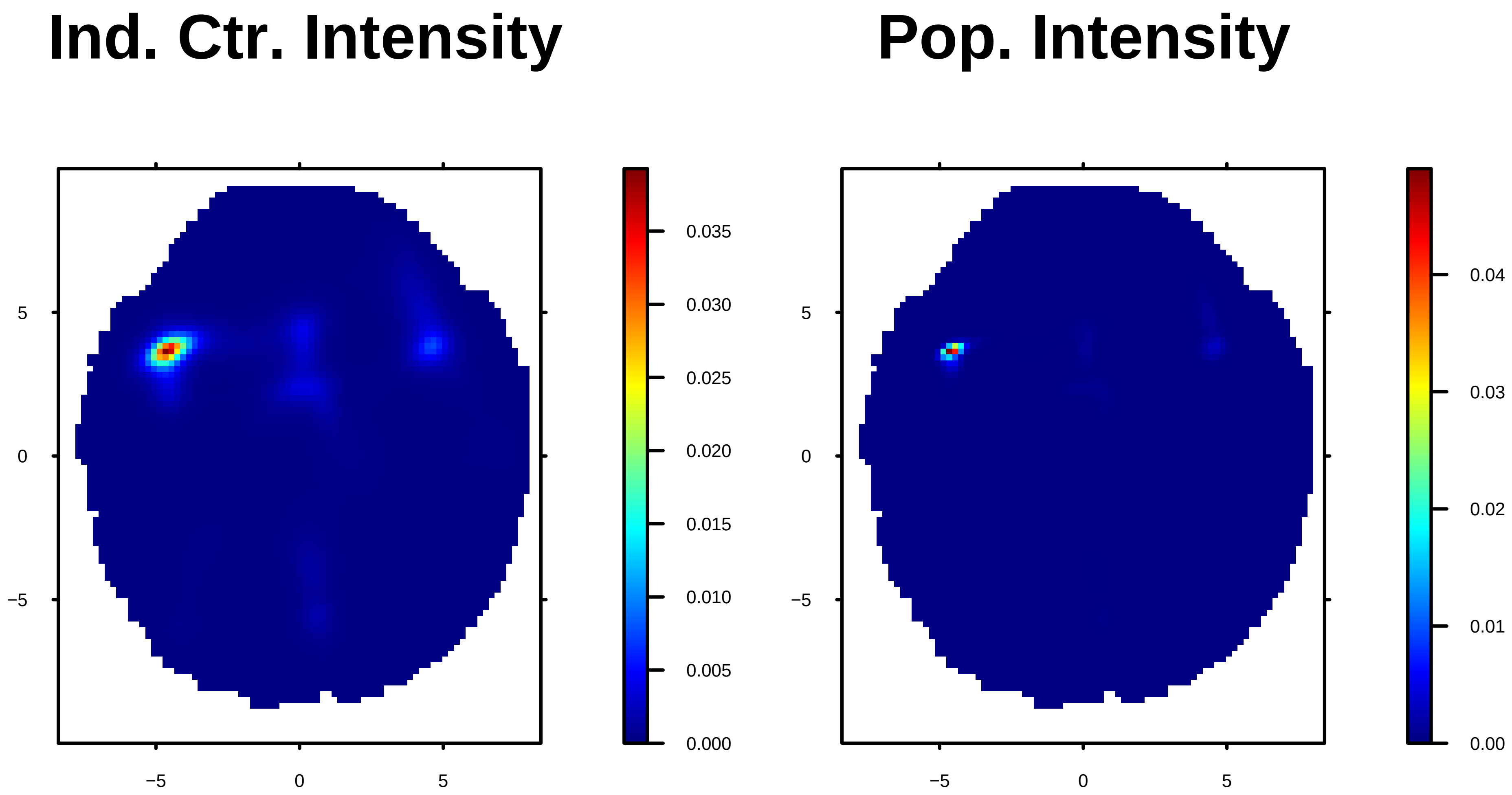


Results: Marginal intensity of Pop Ctrs

$$E(N) = 5.2$$

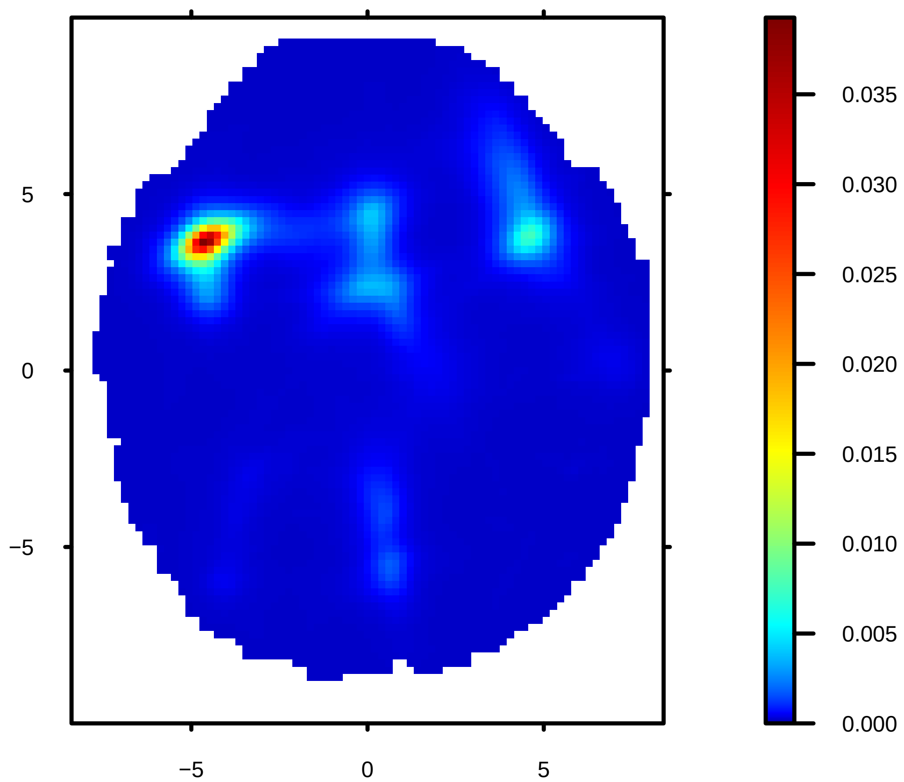


Results: Intensity Functions at Slice 40

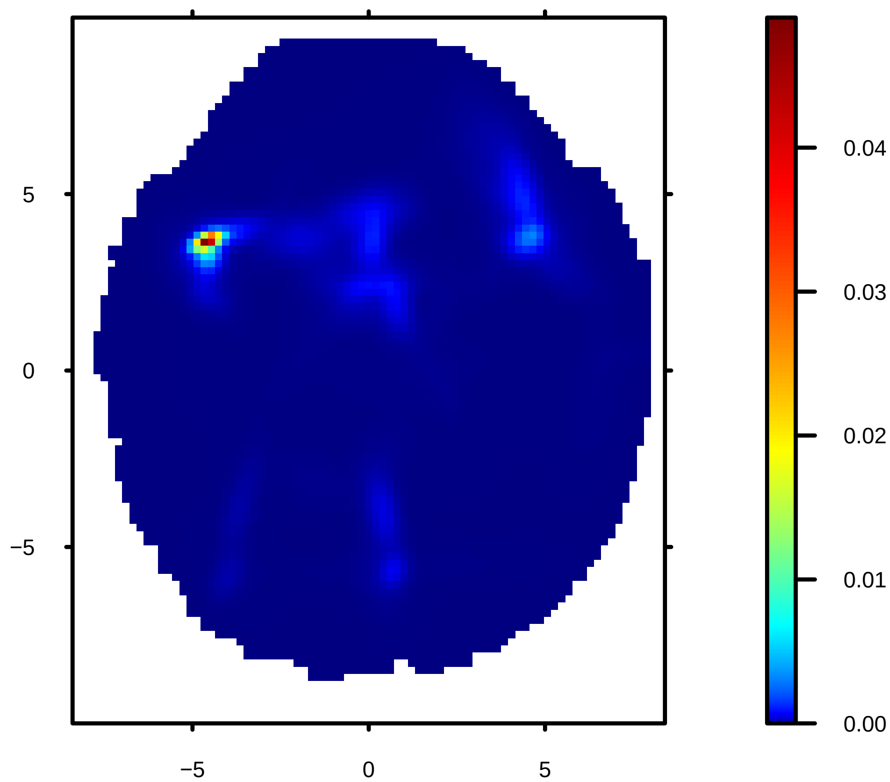


Results: Intensity Functions at Slice 40 (sqrt)

Ind. Ctr. Intensity

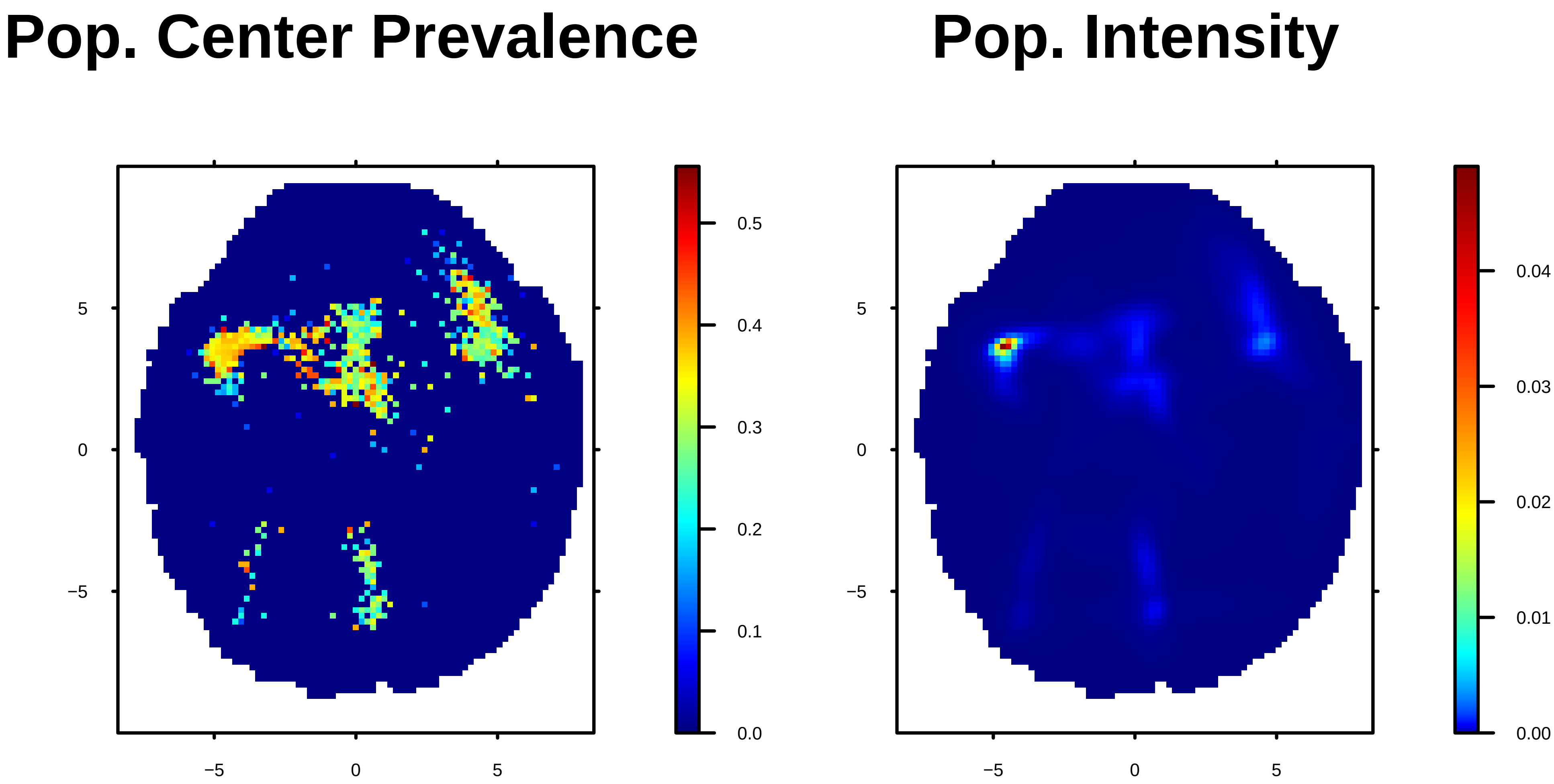


Pop. Intensity



expected number of pop centers in 1 cm^3 cube centered at (-4.6,3.6,1.0) and (4.6,3.6,1.0) is .799 and .136 ,respectively

Results: Population Center Prevalence



Conclusion

- We've shown how a spatial Cox cluster model
  - can be used to quantify the location and spread of population centers
  - can be used to quantify the spread of individual activation centers about population centers
  - ignores activation centers that do not cluster (spurious activation sites)
  - does not rely on overlap of individual activation regions
    - It is not a voxel-level analysis
- Can easily incorporate other relevant prior information
  - e.g. regional brain information (can exclude activations centers in one region of the brain from clustering with activation centers in a neighboring, yet distinct, region)

Acknowledgements

Collaborators

- Lei Xu — Vanderbilt University, Biostatistics
- Tom Nichols — University of Oxford, FMRIB Centre & GlaxoSmithKline
- Derek Nee — Indiana University, Psychology

Funding

- NIH grant PO1 CA087684
- NIH grant RO1 MH069326