Functional Bayesian point process model for neuroimaging meta-analysis data

Silvia Montagna

Joint work with T. D. Johnson and T. E. Nichols

November 27, 2014

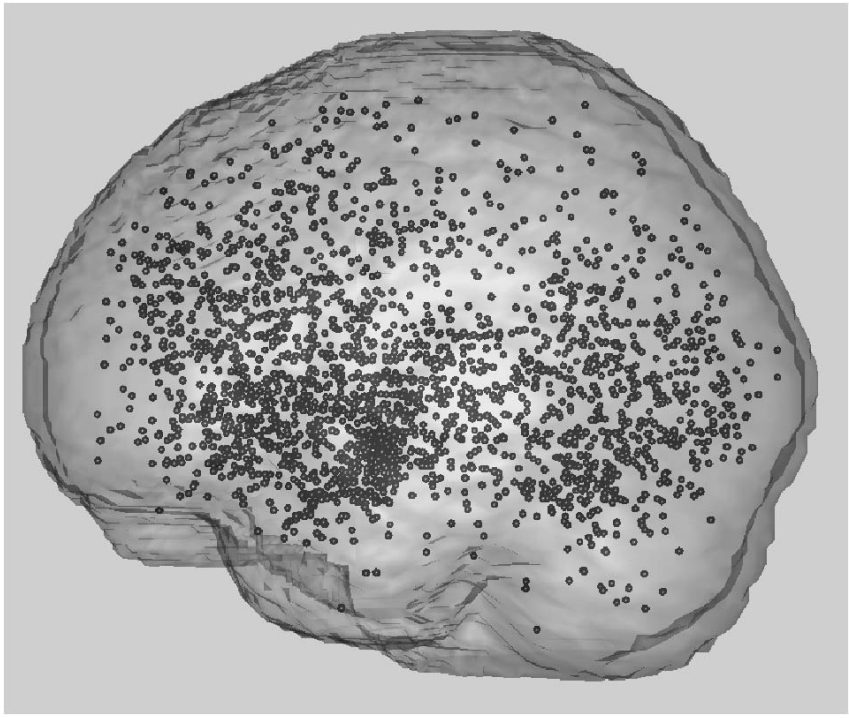
Functional neuroimaging: an overview

- Functional neuroimaging has become an essential tool for non-invasively studying the brain of normal and clinical populations
  - Ex: compare activations between two similar but different types of experiments
  - Increasing interest in inverse inference
- Work supported by the simplicity and computational efficiency of the mass univariate approach
- Limitations: low power, low reliability, etc.
- Need **meta-analyis**!

Coordinate-based meta-analysis

Often, peak activation coordinates (**foci**) only are reported

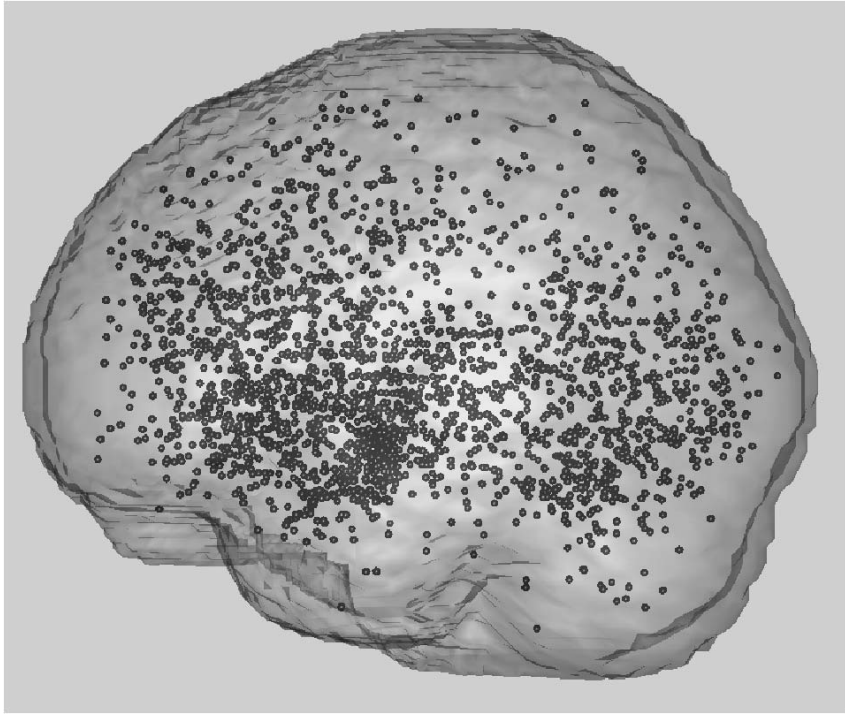
L Fusiform (19/37)	−36	−70	−12
L Fusiform (19/37)	−38	−64	−12
L Inferior temporal (37)	−42	−58	−16
R Cerebellum	4	−66	−24
R Cerebellum	18	−68	−24
R Sup./inf. parietal (19/39/40)	28	−66	36
L Inferior parietal (19/39/40)	−30	−50	36
L Middle temporal (21)			
R Postcentral	36	−26	44
L Postcentral/precentral	−44	−8	36
L Precentral			
R Thalamus	2	−4	0
L Thalamus			
L Lenticular nucleus			
L Paracentral			
L Mid/anterior cingulate	−4	4	44
L Inferior frontal (44)	−42	6	20
L Inferior frontal (44)	−40	10	20



Coordinate-based meta-analysis

Often, peak activation coordinates (**foci**) only are reported

L Fusiform (19/37)	−36	−70	−12
L Fusiform (19/37)	−38	−64	−12
L Inferior temporal (37)	−42	−58	−16
R Cerebellum	4	−66	−24
R Cerebellum	18	−68	−24
R Sup./inf. parietal (19/39/40)	28	−66	36
L Inferior parietal (19/39/40)	−30	−50	36
L Middle temporal (21)			
R Postcentral	36	−26	44
L Postcentral/precentral	−44	−8	36
L Precentral			
R Thalamus	2	−4	0
L Thalamus			
L Lenticular nucleus			
L Paracentral			
L Mid/anterior cingulate	−4	4	44
L Inferior frontal (44)	−42	6	20
L Inferior frontal (44)	−40	10	20



Adopt **spatial point processes**

Goal: build a Bayesian model for coordinate meta-analysis data that allows inference on the most likely domain for any new experiment

Spatial point processes

A spatial **point process** is a stochastic process where we observe random occurrences of events, possibly with a value or mark attached to each occurrence.

Examples:

- Locations of foci in a brain (mark: z- or t-score)
- Locations of trees in a forest (mark: type of tree, diameter, etc.)
- Occurrences of disease (mark: type of disease)

Poisson processes

A **Poisson process X** is defined by a non-negative finite intensity measure M which defines, for any $A \subseteq \mathcal{B}$, the number of expected points in A:

$$N(A) \sim \text{Poisson}(M(A)), \quad \text{where } M(A) \equiv \int_A \mu(\nu) d\nu < \infty$$

We write $\mathbf{X} \sim \text{Poisson}(\mathcal{B}, \mu)$, where $\mathcal{B}$ is a common brain template $\mathcal{B} \subset \mathbb{R}^3$ with finite volume $|\mathcal{B}|$, and μ is the intensity function.

The density of a realisation $\mathbf{x}$ of $\mathbf{X}$ is

$$\pi(\mathbf{x}) = \exp\{|\mathcal{B}| - M(\mathcal{B})\} \prod_{x \in \mathbf{x}} \mu(x)$$

A **Cox process** is a Poisson process with random intensity $\mu(x)$.

Assumptions and notation

- N independent observations, i.e., N independent point patterns arising from N studies
- Foci $\mathbf{x}_i = \{\mathbf{x}_{ij}\}_{j=1}^{n_i}$ from study $i = 1, 2, \ldots, N$
- Each $\mathbf{x}_i$ is a realisation from a Cox process $\mathbf{X}_i$ driven by a random intensity function μ_i

Given that observations are independent, the sampling distribution is

$$\pi(\{\mathbf{x}_i\}_{i=1}^N | \{\mu_i\}_{i=1}^N) \propto \exp\left\{-\sum_{i=1}^N M_i(\mathcal{B})\right\} \prod_{i=1}^N \prod_{\mathbf{x}_{ij} \in \mathbf{x}_i} \mu_i(\mathbf{x}_{ij})$$

$$M_i(\mathcal{B}) = \int_{\mathcal{B}} \mu_i(\boldsymbol{\nu}) d\boldsymbol{\nu} < \infty$$

$\mathcal{B}$ is a common brain template $\mathcal{B} \subset \mathbb{R}^3$ with finite volume $|\mathcal{B}|$.

Log-Gaussian Cox process (LGCP)

One model for Cox processes is an LGCP. Writing the intensity as $\mu(\boldsymbol{\nu}) = \exp\{\mathbf{z}^\top(\boldsymbol{\nu})\boldsymbol{\beta}\}\mu_0(\boldsymbol{\nu})$, with $\mathbf{z}$ denoting a set of covariates, one models

$$\mu_0(\boldsymbol{\nu}) = \exp\{\Psi(\boldsymbol{\nu})\}, \qquad \Psi(\boldsymbol{\nu}) \sim \mathrm{GP}(\varphi, C(\boldsymbol{\nu}, \boldsymbol{\nu}')),$$

with $C(\boldsymbol{\nu}, \boldsymbol{\nu}') = \sigma^2 \exp\{-\phi||\boldsymbol{\nu} - \boldsymbol{\nu}'||\}$.

$\mathbb{E}\mu_0(\boldsymbol{\nu}) = \exp\{\varphi + \sigma^2/2\}$, so a useful specification is to set $\varphi = -\sigma^2/2$ so that $\mathbb{E}\mu_0(\boldsymbol{\nu}) = 1$.

Many decent options for fitting LGCPs: INLA, MCMC, MALA, Hamiltonian MC.

Alternatives to LGCPs

- Look into functional representation of the log intensity function
- Natural computational advantage over a GP
- Functional representations use a linear mean structure to describe the intensity's behaviour whereas GPs use the covariance matrix

Functional Latent Factor Regression Model (LFRM)

We write the (log) intensity as

$$\log \mu_i(\boldsymbol{\nu}) = \sum_{m=1}^{p} b_m(\boldsymbol{\nu})\theta_{im} = \mathbf{b}(\boldsymbol{\nu})^\top \boldsymbol{\theta}_i \tag{1}$$

- $b_m(\boldsymbol{\nu})$ corresponds to the mth basis function evaluated at voxel ν
- Potential choices for $\mathbf{b}(\cdot)^\top$: B-splines, 3D Gaussian kernels
- $\boldsymbol{\theta}_i$ is the vector of study-specific basis function coefficients (**scores**)

Variations between intensities are reflected through the variations in the score vectors.

LFRM

We specify a sparse factor model for the basis coefficients as

$$\boldsymbol{\theta}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i, \qquad \boldsymbol{\zeta}_i \sim \mathrm{N}_p(0, \boldsymbol{\Sigma}) \tag{2}$$

- $\boldsymbol{\Lambda}$ is a $p \times k$ factor loading matrix
- $\boldsymbol{\eta}_i$ is a $k \times 1$ vector of latent factors
- $\boldsymbol{\zeta}_i$ is a residual vector that is uncorrelated with other variables in the model, with $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$

This structure induces a low-dimensional representation of $\log \mu_i$.

LFRM

Information from study-specific covariates $\mathbf{z}_i$ can be incorporated through a simple linear model

$$\boldsymbol{\eta}_i = \boldsymbol{\iota}^\top \mathbf{z}_i + \boldsymbol{\Delta}_i, \quad \boldsymbol{\Delta}_i \sim N_k(0, \boldsymbol{I}) \tag{2}$$

where $\boldsymbol{\iota}$ is a $r \times k$ matrix of unknown coefficients, and r denotes the dimension of $\boldsymbol{\iota}_i$.

LFRM

Despite the simplicity of this linear model, the resulting model on $\log \mu_1, \ldots, \log \mu_n$ allows a very flexible accommodation of covariate information.

Conditionally on $(\{b_m\}_{m=1}^p, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}, \boldsymbol{\iota}, \{\mathbf{z}_i\}_{i=1}^n)$, the log intensities are independent (finite rank) Gaussian processes with covariate dependent mean functions

$$\mathbb{E}[\log \mu_i(\boldsymbol{\nu})] = \sum_{l=1}^{k} \boldsymbol{\iota}_l^\top \mathbf{z}_i \phi_l(\boldsymbol{\nu})$$

and a common covariance function

$$\mathbb{C}\mathrm{ov}\{\log \mu_i(\boldsymbol{\nu}), \log \mu_i(\boldsymbol{\nu}')\} = \sum_{l=1}^{k} \phi_l(\boldsymbol{\nu})\phi_l(\boldsymbol{\nu}') + \sum_{m=1}^{p} \sigma_m^2 b_m(\boldsymbol{\nu}) b_m(\boldsymbol{\nu}'),$$

where $\phi_l(\boldsymbol{\nu}) = \sum_{m=1}^{p} \lambda_{lm} b_m(\boldsymbol{\nu})$.

Bayesian analysis: the MGPS prior

We adopt the multiplicative gamma process shrinkage (MGPS) prior on the loadings [1]:

$$\begin{aligned} \lambda_{jh}|\psi_{jh},\tau_h &\sim N(0,\psi_{jh}^{-1}\tau_h^{-1}), \quad \psi_{jh} \sim \text{Gamma}\left(\frac{\rho}{2},\frac{\rho}{2}\right), \quad \tau_h = \prod_{l=1}^{h}\delta_l \\ \delta_1 &\sim \text{Gamma}(a_1,1), \quad \delta_l \sim \text{Gamma}(a_2,1), \quad l \geq 2 \end{aligned}$$

- τ_h is a global shrinkage parameter for the h*th* column
- τ_h's are stochastically increasing under the restriction $a_2 > 1$
- Favors more shrinkage overall as the column index increases
- ψ_{jh}'s are local shrinkage parameters for the elements in the h*th* column
- Avoids over-shrinking the non-zero loadings

MGPS prior

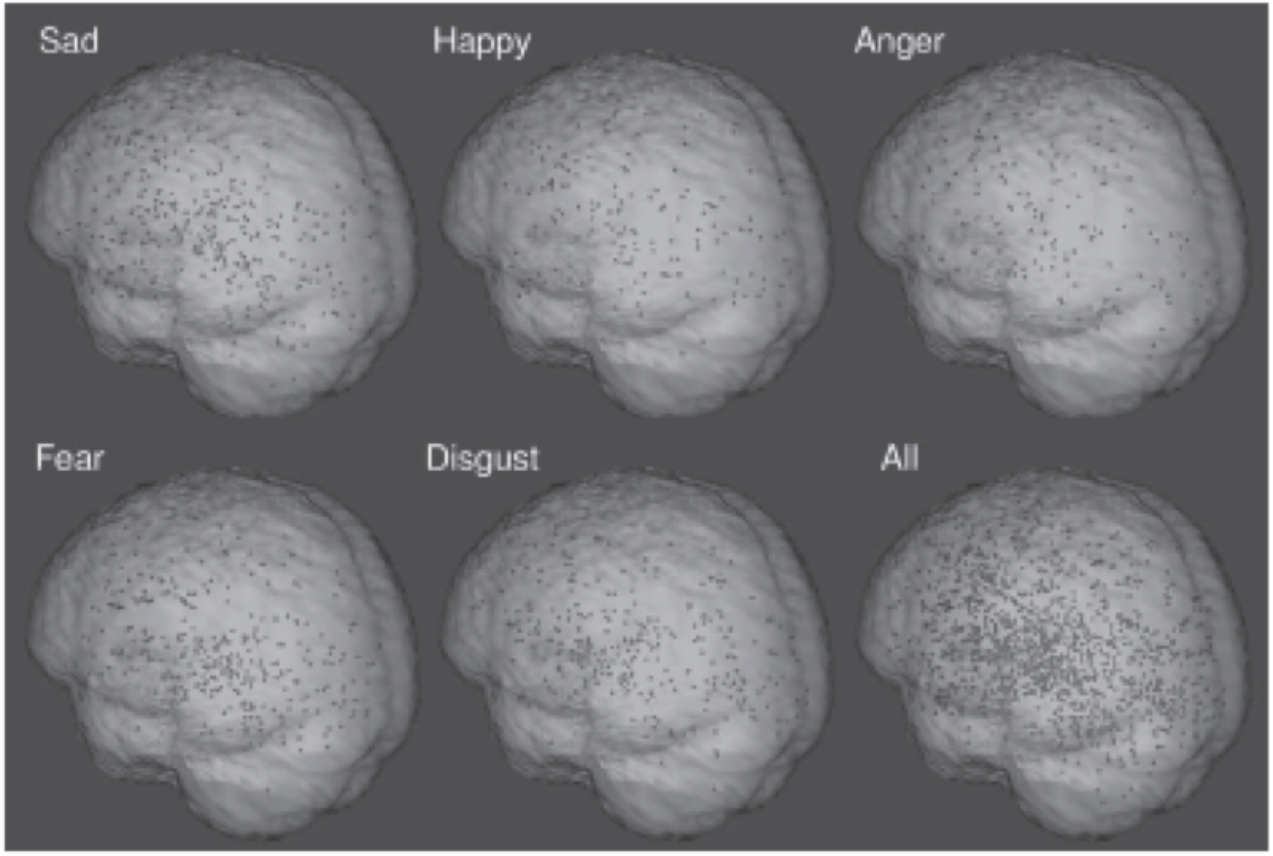
- The MGPS prior let $\mathbf{\Lambda}$ be $q \times \infty$ with priors increasingly concentrated at zero as the column index increases
- The number of factors is automatically selected using adaptive Gibbs sampler
- The MGPS prior allows many of the loadings to be close to zero, thus inducing effective basis selection
- Bhattacharya and Dunson approach - computationally very efficient (block updating)

Posterior computation

- Update the model parameters sampling from their full conditional posterior distributions when available in closed form
- Resort to Hamiltonian MC to update the basis coefficients θ_i
- Adapt the number of factors as the sampler progresses

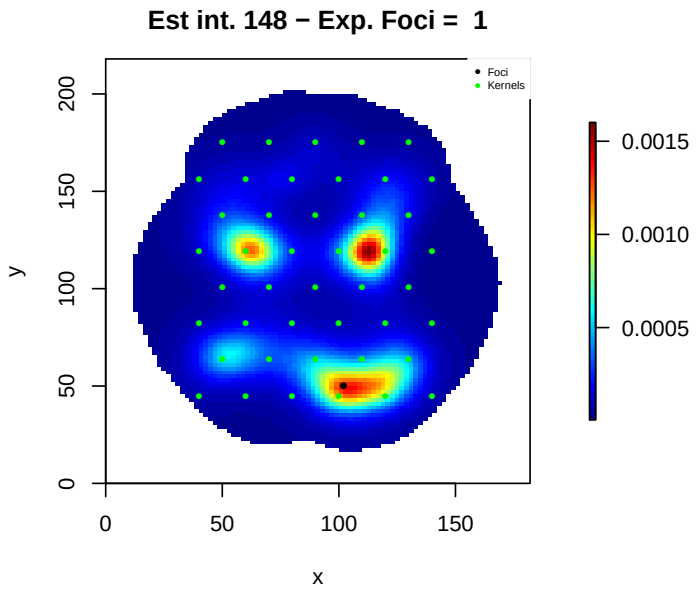
Emotion meta-analysis dataset

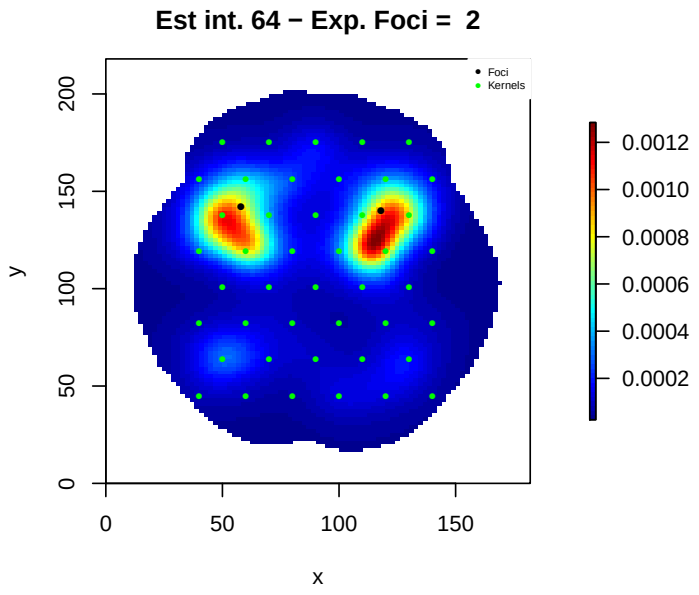
- 164 publications
- 219 studies and 1393 foci
- Five emotions: sad, happy, anger, fear, and disgust

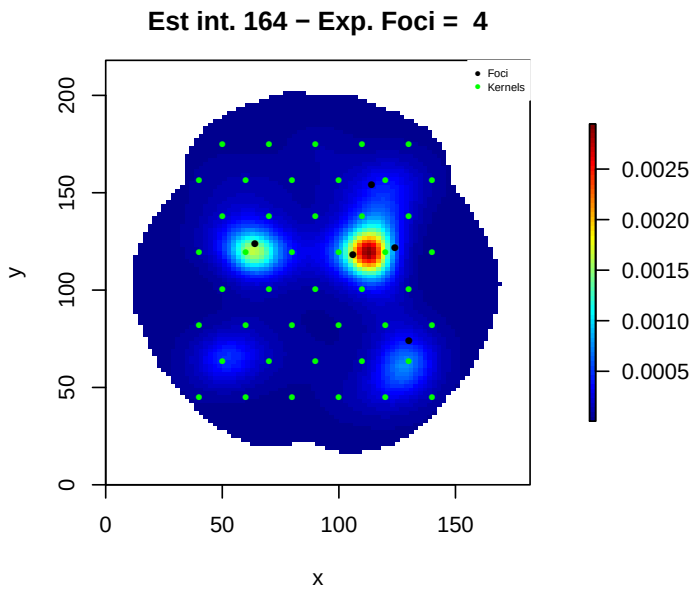


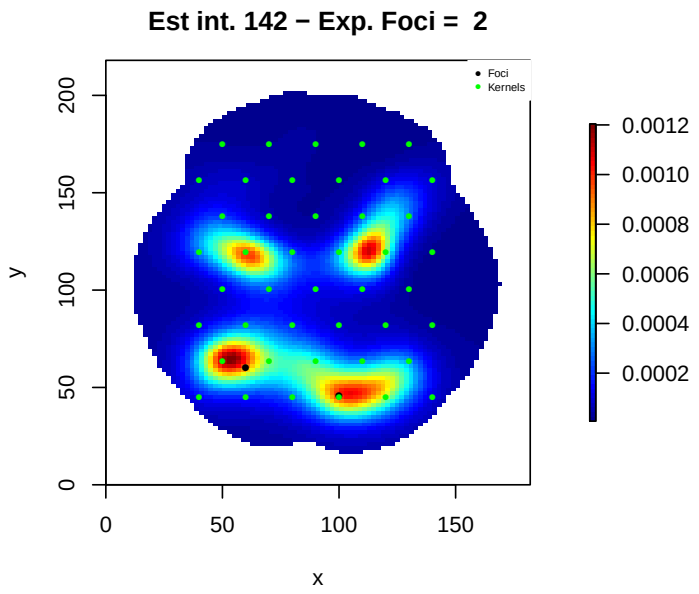
(Picture from [2])

Some estimated intensities

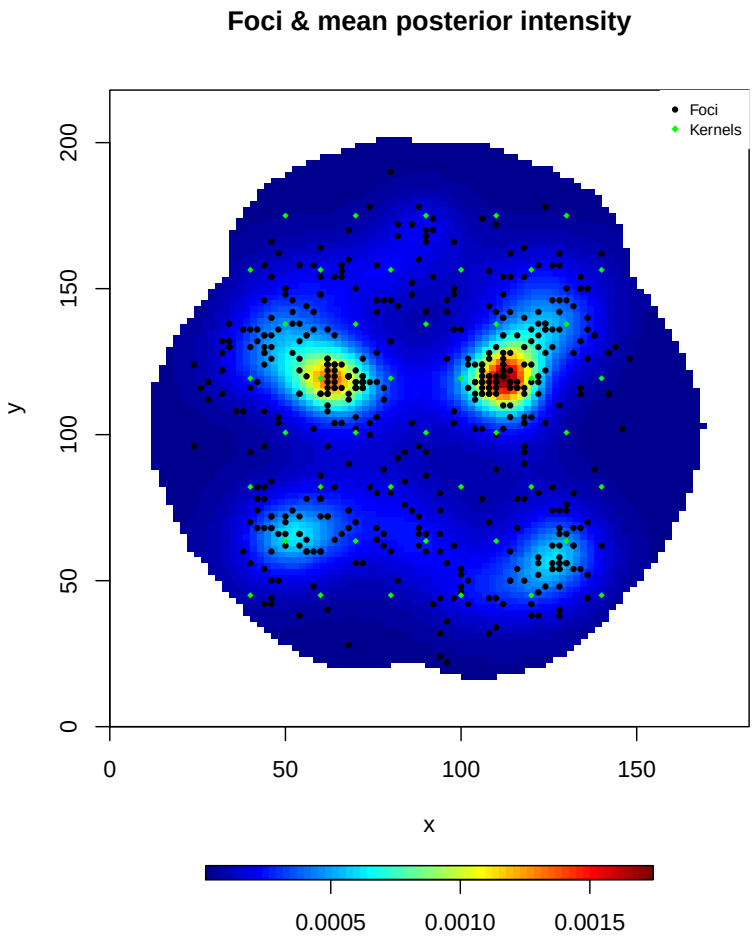


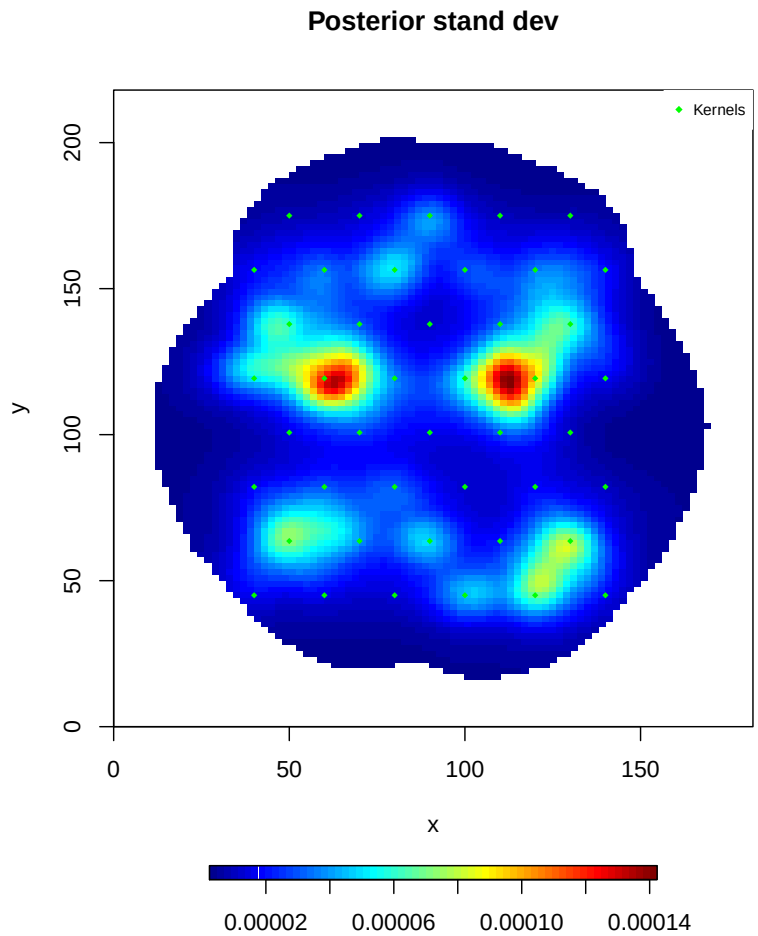






Summary


Foci & mean posterior intensity
Foci
Kernels
y
x
0
50
100
150
200
0.0005
0.0010
0.0015


Posterior stand dev
Kernels
y
x
0
50
100
150
200
0.00002
0.00006
0.00010
0.00014

Probit model for valence

Consider a meta-analysis data set of emotion studies, which can be split into positive and negative studies by valence. Let

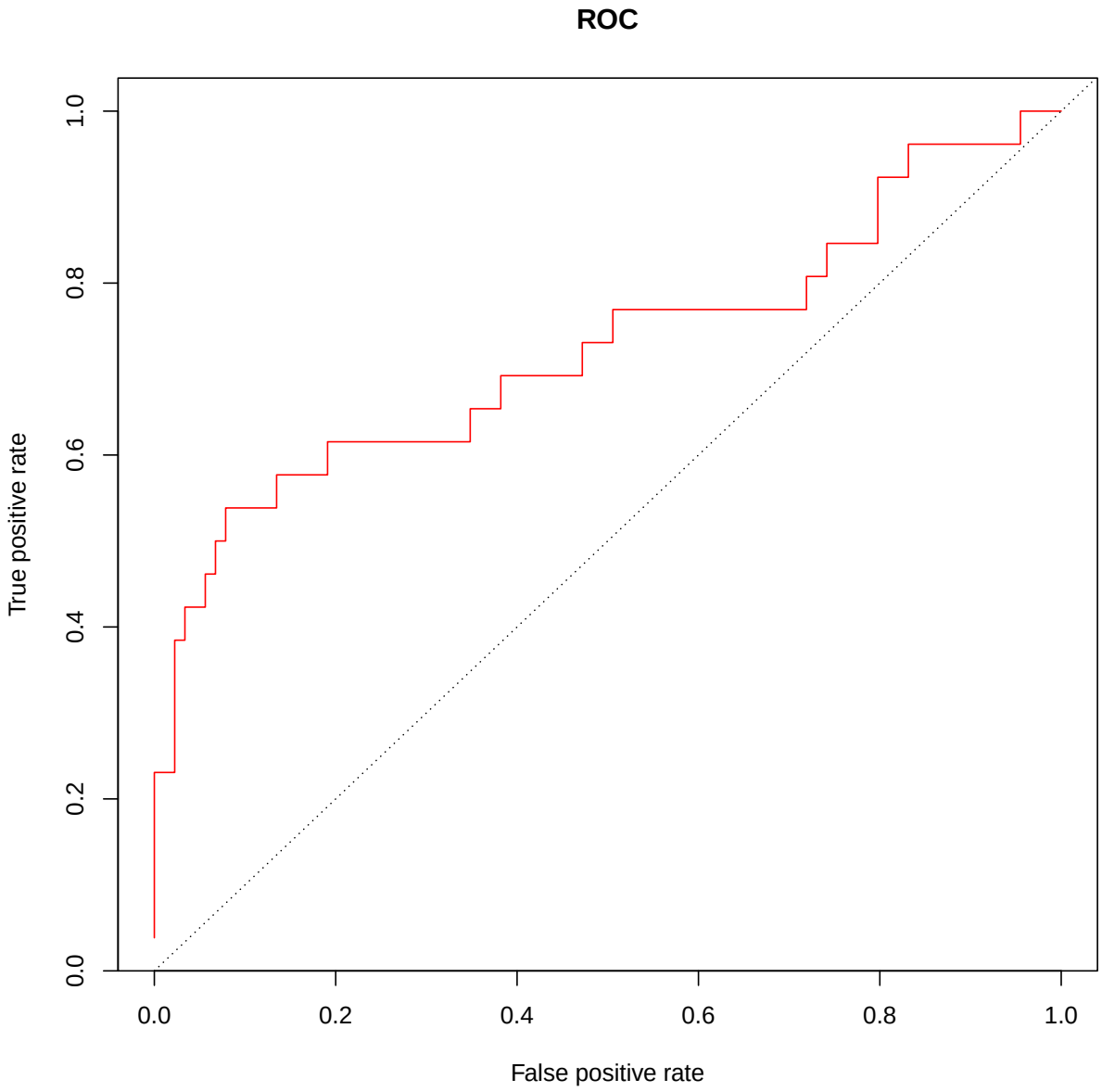
$$y_i = \begin{cases} 1 & \text{if study } i \text{ is positive} \\ 0 & \text{if study } i \text{ is negative} \end{cases}$$

with $\mathrm{P}(y_i = 1 | \alpha, \gamma, \eta_i) = \Phi(\alpha + \gamma^\top \eta_i)$.

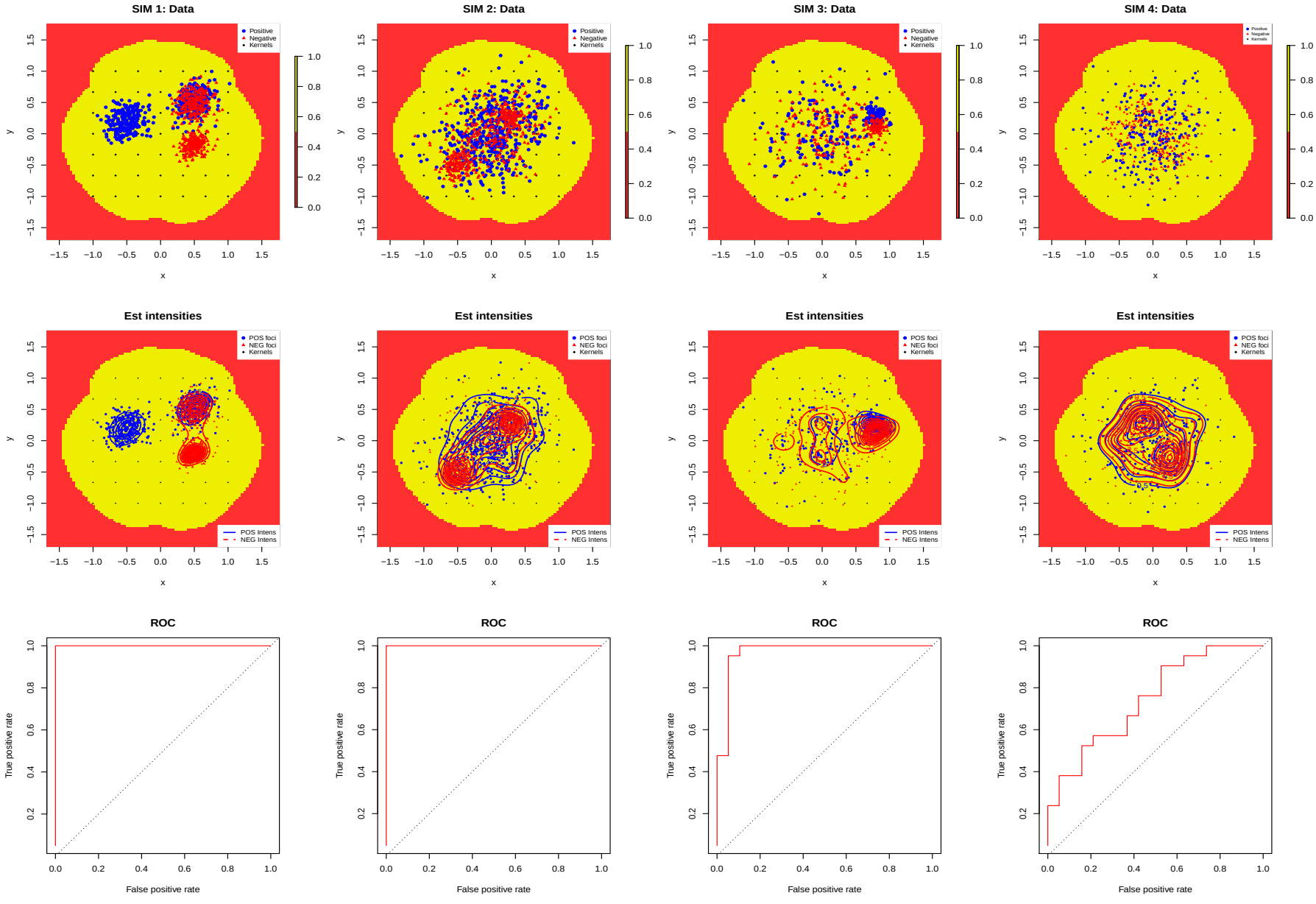
The same set of latent factors impacts on the intensity function via the basis coefficients θ_i and on the response variable via the probability of a positive study.

The model is extendable to different types of outcomes.

ROC – Emotion dataset



Simulations



Summary and Discussion

- Functional representation of the log intensity function of a Cox process with inclusion of a high-dimensional set of pre-specified basis functions
- Allow for automatic shrinkage and effective removal of basis coefficients that are not needed to characterise the intensities
- Covariates allowed to impact on the latent factor scores
- Easy modifications for joint modelling of disparate data of many different types
- θ_i replaceable with concatenated coefficients within component models for different types of objects, including images, movies, text, etc.
- Extendable to a semiparametric case that allows the latent variables densities to be unknown via nonparametric Bayes priors

References

Anirban Bhattacharya and David B Dunson.

Sparse Bayesian infinite factor models.
Biometrika, 98(2):291–306, June 2011.

Jian Kang, Timothy D Johnson, and Thomas E Nichols.

A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis.
Annals of Applied Statistics, 8(3):1800–1824, 2014.