Spatial Bayesian latent factor regression modeling of coordinate-based meta-analysis data

(Article begins on next page)

21 November 2023

Spatial Bayesian Latent Factor Regression Modeling of Coordinate-based Meta-analysis Data

Silvia Montagna[*1], Tor Wager[2], Lisa Feldman-Barrett[3], Timothy D. Johnson[4] and Thomas E. Nichols[1]

[1]Department of Statistics, University of Warwick, CV4 7AL Coventry, UK

[2]Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, CO 80309, U.S.A.

[3]Department of Psychology, Northeastern University, Boston, MA 02115, U.S.A.

[4]Biostatistics Department, University of Michigan, Ann Arbor, MI 48109, U.S.A.

November 16, 2021

Abstract

Now over 20 years old, functional MRI (fMRI) has a large and growing literature that is best synthesised with meta-analytic tools. As most authors do not share image data, only the peak activation coordinates (foci) reported in the paper are available for Coordinate-based Meta-analysis (CBMA). Neuroimaging meta-analysis is used to 1) identify areas of consistent activation; and 2) build a predictive model of task type or cognitive process for new studies (reverse inference). To simultaneously address these aims, we propose a Bayesian point process hierarchical model for CBMA. We model the foci from each study as a doubly stochastic Poisson process, where the study-specific log intensity function is characterised as a linear combination of a high-dimensional basis set. A sparse representation of the intensities is guaranteed through latent factor modeling of the basis coefficients. Within our framework, it is also possible to account for the effect of study-level covariates (meta-regression), significantly expanding the capabilities of the current neuroimaging meta-analysis methods available. We apply our methodology to synthetic data and a neuroimaging meta-analysis dataset.

---

[*]S.Montagna@warwick.ac.uk

1

1 Introduction

Functional magnetic resonance imaging (fMRI) has become an essential, non-invasive, tool for learning patterns of activation in the working human brain (e.g., Pekar (2006), Wager et al. (2015)). Whenever a brain region is engaged in a particular task, there is an increased demand for oxygen in that region which is met by a localised increase in blood flow. The MRI scanner captures such changes in local oxygenation via a mechanism called the *Blood Oxygenation Level-Dependent* (BOLD) effect; see, e.g., Brown et al. (2007) for a brief introduction on fMRI. The great popularity that fMRI has achieved in recent years is supported by various software packages that implement computationally efficient analysis through a mass univariate approach (MUA). Specifically, MUA consists of fitting a general linear regression model at each voxel independently of every other voxel, thus producing images of parameter estimates and test statistics. These images are then thresholded to identify significant voxels or clusters of voxels, and significance is determined via random field theory (Worsley et al. 1996) or permutation methods (Nichols & Holmes 2001). Despite its simplicity, the MUA lacks an explicit spatial model. Even though the activation of nearby voxels is correlated estimation with the MUA ignores the spatial correlation, but inference later accounts for it when random field theory or permutation procedures define a threshold for significant activation.

The relatively high cost of MRI scanner time, however, pose some limitations to single fMRI studies. The main limitation is the small number of subjects that can be recruited for the study, often fewer than 20 (Carp 2012). As a result, most fMRI studies suffer from inflated type II errors (i.e., low power) and poor reproducibility (Thirion et al. 2007). To overcome these limitations there has been an increasing interest in the meta-analysis (MA) of neuroimaging studies. By combining the results of independently conducted studies, MA increases power and can be used to identify areas of consistent activation while discounting chance findings.

In addition to the identification of areas of consistent activation (forward inference, or Pr[Activation | Task]), there has been intense interest in the development of meta-analytic methods to implement proper reverse inference (Yarkoni et al. 2011). Reverse inference refers to inferring which cognitive process or task generated an observed activation in a certain brain region (Pr[Task | Activation]).

2

Suppose that researchers develop a task to probe cognitive process A and find that brain area X is activated. A common but misguided practice in neuroimaging is to conclude that activation of brain region X is evidence that cognitive process A is engaged. However, this logic is wrong and the resulting inference is faulty. In fact, it is known that a single region may be activated by several different tasks (Yeo et al. 2015).

Given that published fMRI studies rarely share the statistic images or raw data, MA techniques are typically based on coordinates of activation, that is, the (x, y, z) coordinates of local maxima in significant regions of activation, where the coordinate space is defined by a standard anatomical atlas. We shall refer to these coordinates as foci (singular focus), and denote the MA based on foci as Coordinate-Based Meta-Analysis (CBMA). Several approaches to CBMA can be found in the literature. See, for example, Fox et al. (1997), Nielsen & Hansen (2002), Turkeltaub et al. (2002), Wager et al. (2004), Kober et al. (2008), Eickhoff et al. (2009), Radua & Mataix-Cols (2009), Kang et al. (2011), Yue et al. (2012), Kang et al. (2014). These methods can be categorised as either kernel-based or model-based approaches.

The most popular kernel-based approaches to CBMA are activation likelihood estimation (Turkeltaub et al. (2002), ALE), modified ALE (Eickhoff et al. (2009), modALE), and multilevel kernel density analysis (Wager et al. (2007), Kober et al. (2008), MKDA). These methods proceed in three main steps. First, one creates *focus maps* for each focus in each study; in these images the intensity at each voxel depends on the proximity of that voxel to that map's focus. For each study, there are as many focus maps as the number of reported foci. These focus maps are then combined to create study maps, which are further combined into a single statistic image (MA map) that represents the evidence for consistent activation (clustering). Significance of the statistic image is assessed with a Monte Carlo test under the null hypothesis of complete spatial randomness. The difference across the aforementioned methods lie in how they create the foci maps, and in how these maps are combined into study and MA maps. We defer to Samartsidis et al. (2016) for an extensive review of these methods. These approaches, however, have some serious limitations. In particular, they are based on a MUA, and require ad-hoc spatial kernel parameters that need to be pre-specified.

Recently, model-based approaches have been proposed to overcome some of the limitations of kernel-based methods. See Samartsidis et al. (2016) for an extensive review of model-based methods. All of these methods are grounded in the spatial statistics literature and utilize spatial

stochastic models for the analysis of the foci. However, there are relatively few works that take this approach. Kang et al. (2011) propose a Bayesian spatial hierarchical model using a marked independent cluster process. Despite its flexibility, the model involves many hyperprior distributions whose parameters are challenging to specify and require expert opinion, and the posterior intensity function is somewhat sensitive to the choice of hyperpriors. Yue et al. (2012) propose a Bayesian spatial binary regression model where the probability that a voxel is reported as a focus, $p(\boldsymbol{\nu})$, is modeled via logistic regression, $p(\boldsymbol{\nu}) = \Phi(z(\boldsymbol{\nu}))$, and $z(\boldsymbol{\nu})$ is modeled as a spatially adaptive Gaussian random field. This method, however, does not treat the number and the location of the foci for each study as random. Also, it does not treat the MA studies as the units of observation, but rather the data at each voxel. Further, both Kang et al. (2011) and Yue et al. (2012) propose models for a single, homogeneous group of studies whereas it is common practice in MA to simultaneously consider several types of tasks. To address this limitation, Kang et al. (2014) propose a Bayesian hierarchical Poisson/gamma random field model for multi-type MA by generalizing the Poisson/gamma random field model developed by Wolpert & Ickstadt (1998). The foci from each type of study are modeled as a cluster process driven by a random intensity function that is modeled as a kernel convolution of a gamma random field. The type-specific gamma random fields are related and modeled as a realization of a common gamma random field shared by all types, inducing correlation between study types. Also, the authors propose a model-based classifier to perform reverse inference.

In this paper, we propose a Bayesian hierarchical model that is both simpler and more flexible than previous Bayesian point process models for CBMA (Kang et al. 2011, 2014). In particular, we extend the Bayesian latent factor regression model for longitudinal data of Montagna et al. (2012) to the analysis of CBMA data. We model the foci from each study as a "doubly stochastic" Poisson process (Cox process) (Møller & Waagepetersen 2004), where the study-specific log intensity function is characterised as a linear combination of a 3-dimensional basis set. We adaptively drop unnecessary bases by imposing sparsity on their coefficients through a latent factor regression model, and information on covariates is incorporated through a simple linear regression model on the latent factors. By interpreting the latent factors as vectors of coefficients, our construction becomes analogous to a functional principal component analysis representation of the log intensities, but in our construction bases are no longer mutually orthogonal functions. The number of latent

4

bases is estimated via posterior computation. The latent factorisation is used as a vehicle to link the intensity functions to a study-type as part of a scalar-on-image regression. Specifically, suppose the MA dataset can be split into studies of two types, A and B. We build a probit regression model to predict the study type, where the probability that study i is of type A is expressed as a function of the latent factors. Along the same lines, we can easily accommodate prediction of two or more types of studies. Our fully Bayesian CBMA model permits explicit calculation of a posterior predictive distribution for study type and, as a result, allows inference on the most likely domain for any new experiment by just using its foci.

To demonstrate the application of our algorithm to real-world data, we compared studies of emotion and cognitive control, using hand-coded activation coordinates from previous meta-analyses (Kim et al. 2012, Kober et al. 2008, Lindquist et al. 2012, Nee et al. 2007, 2013, Rottschy et al. 2012, Wager & Smith 2003, Wager et al. 2004). Each domain has been studied extensively using neuroimaging in hundreds of published studies. There is substantial convergence about the systems broadly involved in each, and though they interact, cognitive control and emotion are associated with distinct large-scale networks. In addition, there is substantial converging evidence on the cognitive and emotional functions of homologous systems in invasive animal studies, further validating the functional roles of the systems identified in human neuroimaging. We aimed to both characterize the patterns of brain activation that are typical to each type, as well as evaluate the ability of our model to conduct a limited "reverse inference", that is, to predict which topic (cognitive control or emotion) a new study investigates.

The remainder of this paper is organised as follows. Section 2 describes our spatial latent factor regression model for CBMA data and outlines a connection with fPCA. In Section 3, we apply our model to the MA dataset of emotion and executive control studies, and compare our results with MKDA. Section 4 reports results from sensitivity analyses and simulation studies. We conclude the manuscript with a final discussion of our model (Section 5). An outline of the MCMC algorithm for posterior computation and an additional simulation study are reported in the Supplementary Materials.

2 Spatial Bayesian latent factor regression for CBMA

In this Section, we present our spatial Bayesian latent factor regression for CBMA data. Articles often report results from different statistical comparisons called contrasts, hereafter called studies. Following the convention of existing neuroimaging CBMA, we treat the studies as independent. The model outlined in Section 2.1 generalizes Montagna et al. (2012) to the case where observations are spatial point patterns from different studies. Each spatial point pattern is assumed to be an independent realization of a spatial point process. In Section 2.2 we show how the model accommodates reverse inference. Section 2.3 further discusses the methodgoloy by presenting an analogy with functional principal component analysis (fPCA). Section 2.4 presents the prior specification for the model parameters.

2.1 The model

Consider independent spatial point patterns arising from n studies, $\mathbf{x}_1, \ldots, \mathbf{x}_n$. We regard $\mathbf{x}_i$ as a realization of a doubly stochastic Poisson process (Cox 1955) $\mathbf{X}_i$ driven by a non-negative random intensity function μ_i defined on a common brain template $\mathcal{B} \subset \mathbb{R}^3$ with finite volume $|\mathcal{B}|$. Given that observations are independent, the sampling distribution is

$$
\begin{aligned}
\pi(\{\mathbf{x}_i\}_{i=1}^n | \{\mu_i\}_{i=1}^n) \quad &\propto \quad \prod_{i=1}^n \left[\exp\left\{ -M_i(\mathcal{B}) \right\} \prod_{\mathbf{x}_{ij} \in \mathbf{x}_i} \mu_i(\mathbf{x}_{ij}) \right] \\
&= \quad \exp\left\{ -\sum_{i=1}^n M_i(\mathcal{B}) \right\} \prod_{i=1}^n \prod_{\mathbf{x}_{ij} \in \mathbf{x}_i} \mu_i(\mathbf{x}_{ij}),
\end{aligned} \tag{1}
$$

where $\mathbf{x}_i = \{\mathbf{x}_{ij}\}_{j=1}^{n_i}$ is the set of foci reported by study i, $\mathbf{x}_{ij} = (x_{ij1}, x_{ij2}, x_{ij3})^\top$ represents the centre of a voxel (or vertex), and $M_i(B)$ denotes a non-negative intensity measure, $M_i(B) = \int_B \mu_i(s)ds < \infty$, for any Borel measurable subset $B \subseteq \mathcal{B}$. To simplify the notation, we will denote a focus in the brain as $\boldsymbol{\nu}$ hereafter.

For the modeling of the random functions $\mu_1, \ldots, \mu_n$, we follow Montagna et al. (2012). Specifically, we consider a functional representation for the (log) intensity function and write $\log \mu_i$ in terms of a collection of basis functions

$$
\log \mu_i(\boldsymbol{\nu}) = \sum_{m=1}^p \theta_{im} b_m(\boldsymbol{\nu}) = \mathbf{b}(\boldsymbol{\nu})^\top \boldsymbol{\theta}_i. \tag{2}
$$

6

This specification implies that the log intensity function belongs to the span of a set of basis functions, $\{b_m(\cdot)\}_{m=1}^p$, with $\boldsymbol{\theta}_i$ denoting a vector of study-specific coefficients. Choosing the functions $\{b_m(\cdot)\}_{m=1}^p$ is particularly challenging since the appropriate basis is not known in advance and, conceptually, any bases can be chosen. For example, the B-spline basis or Gaussian kernels can be used to model smooth μ_i intensities. Hereafter, we use 3D isotropic Gaussian kernels

$$b_1(\boldsymbol{\nu}) = 1, \quad \text{and} \quad b_m(\boldsymbol{\nu}) = \exp\{-b||\boldsymbol{\nu} - \boldsymbol{\psi}_m||^2\}, \quad m = 2,\ldots,p, \tag{3}$$

with kernel locations $\{\boldsymbol{\psi}_m\}_{m=2}^p$ and bandwidth b to be specified according to prior knowledge. More flexible approaches allow the number and locations of the kernels to be unknown and estimated by the sampler, at the expense of a great increase in computational cost. Hereafter, we prefer adopting a computational-savvy approach by fixing the bases, and use sensitivity analysis to help us determine reasonable choices for p and kernel locations.

The functional representation considered in (2) constitutes an alternative to the typical log Gaussian Cox process prior on μ_i (LGCP, Møller et al. (1998)), which is a widely popular prior within the spatial statistics literature. As its name suggests, the LGCP is a Cox process with $\mu_i(\boldsymbol{\nu}) = \exp\{Z(\boldsymbol{\nu})\}$, where Z is modeled as a Gaussian process. The most attractive feature of this model is that it provides a flexible and relatively tractable construction for describing spatial phenomena. Inference for LGCPs is, however, a computationally challenging problem, and the main barrier is the computation of the covariance matrix of Z. In a typical neuroimaging application, this matrix is very large as its dimensions correspond to the number of voxels in the brain mask (typically, more than 150,000 voxels on a $2 \times 2 \times 2$ mask). Fortunately, for covariance functions defined on regular spatial grids there exist fast methods for computing the covariance based on the discrete Fourier transform (Wood & Chan 1994, Rue & Held 2005, Taylor & Diggle 2014). A basis function representation as in (2) completely removes the need of computing the covariance matrix (and its inverse), hence has a natural computational advantage over LGCPs in this regard.

By characterising the study-specific log intensity functions by a vector of coefficients with respect to a common functional basis representation, all variation between the study-specific intensities are reflected through the variation in the vectors $\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_n$. Montagna et al. (2012) remark that a major drawback of the basis function approach is, however, the failure to obtain a low dimensional representation of the individual intensities. Low dimensional representations are crucial when building a hierarchical model where the foci are to be linked, as predictors or outcomes, with

7

other variables under study. In our construction, the μ_i's are represented by the long vector of coefficients $(\theta_{i1}, \ldots, \theta_{ip})$. Unless the μ_i's are sparse in the chosen basis, these vectors are dense, meaning that any projection of these vectors onto a lower dimensional space results in a substantial loss of information. To obtain a low-dimensional representation of $\log \mu_i$, we follow the lead in Montagna et al. (2012) and place a sparse latent factor model (Arminger & Muthén 1998) on the basis coefficients

$$\boldsymbol{\theta}_i = \boldsymbol{\Lambda}\boldsymbol{\eta}_i + \boldsymbol{\zeta}_i, \quad \text{with} \quad \boldsymbol{\zeta}_i \sim N_p(0, \boldsymbol{\Sigma}) \tag{4}$$

where $\boldsymbol{\theta}_i = [\theta_{i1}, \ldots, \theta_{ip}]^\top$, $\boldsymbol{\Lambda}$ is a $p \times k$ factor loading matrix with $k \ll p$, $\boldsymbol{\eta}_i = (\eta_{i1}, \ldots, \eta_{1k})^\top$ is a vector of latent factors for study i, and $\boldsymbol{\zeta}_i = (\zeta_{i1}, \ldots, \zeta_{ip})^\top$ is a residual vector that is independent with the other variables in the model and is normally distributed with mean zero and diagonal covariance matrix $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_p^2)$. Vectors $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n$ can be put in any flexible joint model with other variables of interest. For example, information from covariates $\mathbf{Z}_i$ can be incorporated through a simple linear model

$$\boldsymbol{\eta}_i = \boldsymbol{\beta}^\top \mathbf{Z}_i + \boldsymbol{\Delta}_i, \quad \text{with} \quad \boldsymbol{\Delta}_i \sim N_k(0, \boldsymbol{I}), \tag{5}$$

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

Despite the simplicity of this hierarchical linear model, the resulting structure on $\log \mu_i(\boldsymbol{\nu})$ allows a very flexible accommodation of covariate information. Specifically, if we marginalise out $\{\boldsymbol{\theta}_i, \boldsymbol{\eta}_i\}$, our model results in a (finite rank) GP for $\log \mu_i$ with covariate dependent mean function

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

and common covariance function

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

where $\phi_l(\boldsymbol{\nu}) = \sum_{m=1}^{p} \lambda_{lm} b_m(\boldsymbol{\nu})$, and $\boldsymbol{\beta}_l$ is the lth column of $\boldsymbol{\beta}$.

2.2 Reverse Inference

In response to an increasing interest in reverse inference, we focus on the development of a methodology which accommodates joint modeling of neuroimaging point pattern data with study types. Suppose we have new point pattern data $\mathbf{x}_{new}$ that is a realization from one of T tasks or cognitive

processes, y_{new}. Further, we have point pattern data from n studies for which the corresponding task or cognitive process is known, $\{\mathbf{x}_i|y_i\}_{i=1}^n$ with $y_i \in \{1,\ldots,T\}$. Interest is in quantifying the probability that the new point pattern data arose from a specific task type, that is, the posterior predictive probability that $\mathbf{x}_{new}$ originates from type t, $\Pr(y_{new} = t|\mathbf{x}_{new}, \{\mathbf{x}_i|y_i\}_{i=1}^n)$. Our fully Bayesian model for neuroimaging point pattern data allows inference on the most likely domain for any new experiment.

Hereafter, we extend the model with focus on our motivating application. The MA dataset consists of studies that can be categorized as either "emotion" or "executive control" (see Section 3 for details). Therefore, the interest is in estimating the probability that newly observed point pattern data arose from either an emotion or executive control experiment. Let y_i denote the study type, with

$$y_i = \begin{cases} 1 & \text{if study } i \text{ is an emotion study} \\ 0 & \text{if study } i \text{ is an executive control study.} \end{cases}$$

Because the study type can be represented as a binary response, we can build a probit model for study type and predict the posterior probability that a new point pattern data arose from either cognitive process. Specifically, we model $p_{y_i} = \Pr(y_i = 1|\alpha, \boldsymbol{\gamma}, \boldsymbol{\eta}_i) = \Phi(\alpha + \boldsymbol{\gamma}^\top \boldsymbol{\eta}_i)$, where $\Phi(\cdot)$ denotes the standard normal distribution function. Parameter α can be interpreted as the baseline probability that study i is of the emotion type, and $\boldsymbol{\gamma}^\top \boldsymbol{\eta}_i$ accounts for study-specific random deviations. Notice that the latent factors $\boldsymbol{\eta}_i$ (Section 2.1) are used as a vehicle to link the random intensities (thus, the foci) to the study-type.

The intercept α is given a $\mathrm{N}(m_\alpha, v_\alpha)$ prior, with $m_\alpha = \Phi^{-1}(0.50)$ assuming emotion and executive control studies are equally likely a priori, and $\boldsymbol{\gamma}$ is a vector of unknown regression coefficients with conjugate prior distribution $\boldsymbol{\gamma} \sim \mathrm{N}_k(\boldsymbol{\mu}_\gamma, \boldsymbol{\Sigma}_\gamma)$. The full conditional posterior distributions needed for Gibbs sampling are not automatically available, but we can rely on the data augmentation algorithm of Albert & Chib (1993) to facilitate the computation. We introduce independent unobservable latent variables $W_1, \ldots, W_n$ and define

$$y_i = \mathbb{1}(W_i > 0), \qquad \text{with} \qquad W_i \sim \mathrm{N}(\alpha + \boldsymbol{\gamma}^\top \boldsymbol{\eta}_i, 1), \tag{6}$$

so that $\Pr(y_i = 1|\alpha, \boldsymbol{\gamma}, \boldsymbol{\eta}_i) = \Phi(\alpha + \boldsymbol{\gamma}^\top \boldsymbol{\eta}_i)$ by marginalizing out the latent variable W_i.

Although the presented is for 2-class CBMA data, the proposed framework can be easily modified for joint modeling of data of many different types, e. g., the probit model for a binary outcome

can be replaced by an appropriate predictive model for categorical, nominal, or continuous study features. The key idea is to use the low dimensional vectors $\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_n$ in all subsequent parts of the model where one seeks to link intensities $\log \mu_1, \ldots, \log \mu_n$ with other variables of interest.

We close this Section by providing a graphical representation (Figure 1) of the spatial Bayesian latent factor model outlined above and in § 2.1. The vector of latent factors plays the key role in linking the two component models for study-type and random intensities, and the study-type y_i is conditionally independent of all nodes in the random intensity model given the latent factors. All parameters located outside of the dashed rectangle $(\alpha, \boldsymbol{\gamma}, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma})$ are shared and estimated by pulling information across all studies, thus allowing for borrowing of information. If covariate information is available, covariates impact on the $\boldsymbol{\eta}_i$'s via a linear regression model. In our motivating application, however, no covariates are available and the covariates component box is omitted by simply considering a standard multivariate normal prior on $\boldsymbol{\eta}_i$.

[Figure 1 about here.]

2.3 fPCA-analogue construction

The vector of latent factors $\boldsymbol{\eta}_i$ can also be interpreted as a coefficient vector by writing

$$\log \mu_i(\boldsymbol{\nu}) = \sum_{l=1}^{k} \eta_{il} \phi_l(\boldsymbol{\nu}) + r_i(\boldsymbol{\nu}), \tag{7}$$

with

$$\phi_l(\boldsymbol{\nu}) = \sum_{m=1}^{p} \lambda_{lm} b_m(\boldsymbol{\nu}) \quad \text{and} \quad r_i(\boldsymbol{\nu}) = \sum_{m=1}^{p} \zeta_{im} b_m(\boldsymbol{\nu}), \tag{8}$$

where $\{\phi_l\}_{l=1}^{k}$ forms an unknown non-local basis set to be learnt from the data and r_i is a function-valued random intercept.

We recall that the GP model can be viewed as an infinite dimensional basis-function expansion. For example, the Karhunen-Loéve expansion of a GP f (with known covariance parameters) at $\boldsymbol{\nu}$ can be written as $f(\boldsymbol{\nu}) = \sum_{k=1}^{\infty} w_k e_k(\boldsymbol{\nu})$, where the basis functions e_k are orthogonal and the coefficients $\{w_k\}$ are independent, zero-mean normal random variables. The variance of w_k is equal to the kth largest eigenvalue. The empirical version (i.e., with the coefficients computed from a sample) is known as (functional) principal component analysis. Decomposition (7), without $r_i(\boldsymbol{\nu})$, is analogous to a truncated fPCA representation of $\log \mu_i(\boldsymbol{\nu})$, however bases $\phi_1(\cdot), \ldots, \phi_k(\cdot)$ are

no longer mutually orthogonal within our construction. Although orthogonality enhances interpretability of the elements of the decomposition, this is not a primary concern in our application because we view the latent factorisation only as a vehicle to link the intensities (and, ultimately, the observed foci) with other variables. To highlight this difference with fPCA, we refer to $\{\phi_l\}_{l=1}^k$ as a dictionary.

The size k and the elements of the dictionary depend on how $\boldsymbol{\Lambda}$ is modeled. This discussion is deferred to the next Section.

2.4 Prior elicitation

Careful modeling of $\boldsymbol{\Lambda}$ is crucial in that the factor loading matrix ultimately controls sparsity and basis selection. Following Montagna et al. (2012), we adopt a multiplicative gamma process shrinkage (MGPS) prior on the loadings (Bhattacharya & Dunson 2011):

$$\lambda_{jh}|\iota_{jh}, \tau_h \quad \sim \quad N(0, \iota_{jh}^{-1}\tau_h^{-1}), \quad \iota_{jh} \sim \text{Gamma}\left(\frac{\rho}{2}, \frac{\rho}{2}\right), \quad \tau_h = \prod_{l=1}^h \delta_l \tag{9}$$

$$\delta_1 \quad \sim \quad \text{Gamma}(a_1, 1), \quad \delta_l \sim \text{Gamma}(a_2, 1), \quad l \geq 2 \tag{10}$$

with $j = 1, \ldots, p$ and $h = 1, \ldots, k$. Elements $\{\delta_l\}_{l \geq 1}$ are independent, τ_h is a global shrinkage parameter for the hth column of $\boldsymbol{\Lambda}$ and the ι_{jh}'s are local shrinkage parameters for the elements in the hth column. Under the choice $a_2 > 1$, the τ_h's are stochastically increasing favouring more shrinkage as the column index increases and preventing the factor splitting problem (Bhattacharya & Dunson 2011). The local shrinkage parameters prevent over-shrinking the nonzero loadings. Therefore, the MGPS prior shrinks a subset of the loadings strongly towards zero while retaining a sparse signal. For more details on the properties of this prior, we defer to Bhattacharya & Dunson (2011).

To obviate the need for pre-specifying the number of factors, we follow Bhattacharya & Dunson (2011) and implement an adaptive algorithm for choosing k. The idea behind this sampler is to strike a balance between missing important factors by choosing k too small and wasting computation on an overly conservative value. At iteration t, one monitors the number of columns of $\boldsymbol{\Lambda}$ having all elements within some pre-specified small neighbourhood of zero. If the number of such columns drops to zero, then a column is added to $\boldsymbol{\Lambda}$ by sampling the new loadings from the prior distribution, and otherwise discard the redundant columns in that the contribution of the factors

is negligible. The other parameters are also modified accordingly. By proposing an oversized set of basis functions (2), we can ultimately control the complexity of our model by including or excluding a particular basis (7) based on its contribution to the likelihood of the observed data. To guarantee convergence of the chain, the adaptations are designed to satisfy the diminishing adaptation condition in Theorem 5 of Roberts & Rosenthal (2007).

The Bayesian specification of our LFRM for spatial point pattern data is completed by placing a Gamma prior on the residual precisions, $\sigma_m^{-2} \sim \text{Gamma}(a_\sigma, b_\sigma)$, for $m = 1, \ldots, p$, and, if covariate information is available, a Cauchy prior on the matrix of coefficients $\boldsymbol{\beta}$ as follows

$$\boldsymbol{\beta}_l \sim \text{N}(0, \text{Diag}(w_{lj}^{-1})), \quad w_{lj} \sim \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right) \tag{11}$$

for $l = 1, \ldots k$ and $j = 1, \ldots, r$. A multivariate Normal prior is also a suitable choice.

The posterior computation for our spatial LFRM proceeds via a hybrid Gibbs sampler with a Hamiltonian Monte Carlo (HMC) step (Neal 2010) to update the basis function coefficients $\boldsymbol{\theta}_i$. Details are provided in the Supplementary Materials (Web Appendix A).

3   Neuroimaging MA application

In this Section, we apply our model to neuroimaging MA data. A typical MA dataset consists of several publications of similar, yet different experiments, and many articles report results from different statistical comparisons, here called studies. Our dataset consists of 1199 total studies categorized as either "emotion" (860 studies, 6481 foci; Kober et al. (2008)) or "executive control" (339 studies, 4332 foci; Kim et al. (2012), Nee et al. (2007), Rottschy et al. (2012), Nee et al. (2013)).

We assigned a Gamma(1, 0.3) prior distribution with mean 1/3 to the diagonal elements of $\boldsymbol{\Sigma}^{-1}$. The hyperparametes of the MGPS prior were set to $\rho = 3$, $a_1 = 2.1$, $a_2 = 3.1$. In absence of covariate information, we assigned $\boldsymbol{\eta}_i \sim \text{N}(\boldsymbol{0}, \boldsymbol{I})$ for $i = 1, \ldots, n$. We chose $p = 352$ Gaussian kernels with bandwidth $b = 0.002 \ mm^2$. Kernels were placed on axial slices roughly 8-12 mm apart, at $z = \{-38, -22, -14, -2, 6, 18, 28, 44\}$ mm (about 85% of the foci were located within these axial slices) and, within each slice, were equally spaced by forming a grid of 6×8 knots along the (x, y) direction. We used a standard brain mask with $2 \ mm^3$ voxels and dimensions $91 \times 109 \times 91$. Kernels falling outside this mask were discarded. To update the basis function coefficients via Hamiltonian

Monte Carlo (Neal 2010), we adopted the leapfrog method for L steps and with a stepsize of ϵ. At each iteration of the MCMC sampler, a new value for L was drawn from Poisson(25) and the stepsize was adapted every 10 iterations during burn-in to benchmark an average acceptance rate of 0.65 over the previous 100 iterations in the Metropolis-Hastings step. The sampler was run for 10,000 iterations, with the first 5,000 samples discarded as a burn-in and collecting every $20th$ sample to thin the chain. We assessed convergence of the chain by multiple runs of the algorithm from over-dispersed starting values and visually inspected the differences in the posterior mean intensity function $\mu_i(\boldsymbol{\nu})$ at a variety of voxels and for different studies. The sampler appeared to converge rapidly and mix efficiently.

Figure 2 displays the posterior mean intensity function at three axial slices for two randomly chosen studies. Slices are arranged in columns, the top (bottom) row shows an emotion (executive control) study. The smoothness of the images is due to the choice of the bandwidth parameter. Larger values for b (same p) induce rougher intensities, but results are robust and with the following conclusions unchanged. As opposed to the executive control study, the emotion study shows strong activation in two regions at slice $z = -20$ and for $y > -26$ mm. These regions correspond to the amygdalae; almond-shaped structures in the brain of known importance in emotion processing. The executive control study shows stronger activation in more superior slices with a bilateral pattern. Note how we identify active regions even though a study does not have foci at a particular slice or, for example, both the right and left amygdala are active for study 232 even though only one focus is reported around the right amygdala. This a by-product of the borrowing of information across studies in our Bayesian treatment and an essential feature to adequately estimate the intensities over the whole the brain given the sparsity of foci per study.

[Figure 2 about here.]

The top row in Figure 3 shows the posterior mean difference between the estimated mean group intensity for the emotion and the executive control studies, respectively. The group intensity at iteration t is obtained by averaging the basis function coefficients for studies that belong to the group, that is,

$$\hat{\mu}_g^t(\boldsymbol{\nu}) = \exp\{\mathbf{b}(\boldsymbol{\nu})^\top \hat{\boldsymbol{\theta}}_g^t\}, \qquad \text{with} \qquad \hat{\boldsymbol{\theta}}_g^t = \frac{1}{\text{Card}(g)} \sum_{i \in g} \hat{\boldsymbol{\theta}}_i^t,$$

where $\text{Card}(g)$ is the cardinality of group g. These maps reflect the degree of consistency with which a region is activated by either task. The darker regions in the top row reveal stronger activation of

emotion studies (the amygdalae at slice $z = -20$ mm), whereas the bright regions denote stronger activation of executive control studies. The grey area reveals no difference in activation. We can also obtain standard deviation maps of the difference map as a measure of uncertainty around the point estimate and the corresponding t-maps. The clearest of these maps is slice $z = -20$ mm where one can easily identify the amygdalae as significant regions.

[Figure 3 about here.]

It is also of interest to examine the elements of the dictionary $\{\phi_l\}_{l=1}^{k}$. The interpretation of the dictionary elements has to be done with care in that they do not constitute orthogonal bases as the eigenfunctions in the fPCA literature. However, examining the dictionary is useful to visualize how the model moves away from the fixed isotropic Gaussian kernels and learns a set of dictionary elements that are useful to represent the intensities. The posterior mean number of latent factors is $k = 4$ with 95% credible interval $[3, 6]$. Figure 4 shows the first four elements of the dictionary $\{\phi_l\}_{l=1}^{4}$ (columns) at axial slices $z = -20, 6, 42$ mms (rows). Notice how the magnitude of the learnt bases decreases as k increases, with the first couple of dictionary elements describing the principal patterns of activation and the later elements progressively shrunk toward zero. This effect is made possible by the MGPS prior on the factor loadings. At every axial slice, the first two dictionary elements combined recover the principal patterns of activation we observed in Figures 2-3. Subsequent dictionary elements are harder to interpret and of more marginal effect.

[Figure 4 about here.]

Our dataset is large enough to split the data into a training set (50%), for which both foci and study type are retained for the analysis, and a testing set, for which the foci only are retained, and we test the predictive accuracy of our model. We compare our predictive method to previous work that combines MKDA and a naïve Bayesian classifier (NBC) (Yarkoni et al. 2011). Using the MKDA framework, this method creates a study-specific binary activation map, where a voxel is given a value of 1 if it is within a 10 mm (Euclidian) distance of a reported focus, and 0 otherwise. For each group (study type), an activation probability map is constructed by taking a weighted average of the binary maps of the studies in that group. Further, the predictive probability of the study type given activation from a new study is then computed using the activation probability maps via Bayes' theorem and under the assumption of independence across voxels. This method is

very computationally efficient, but ignores the spatial dependence in the activation maps, leading to biased predictive probabilities of the class membership. Figure 5 shows the ROC for predicting the study type for study in the test set with corresponding area under the curve (AUC). We see that our predictive model does a better job that MKDA + NBC at predicting the emotion type. This result is robust and we could confirm it through additional chains run for sensitivity analysis to our choice for the model parameters. Therefore, taking into account the spatial information in the data helps in achieving better predictive accuracy, and our Bayesian model captures more sources of variation and conveys the uncertainty in the computation of the predictive probabilities for study type.

[Figure 5 about here.]

4 Simulations and sensitivity analysis

In this Section, we conduct simulation studies to examine how the posterior inference on the intensity functions varies with different specifications of the number of bases p and bandwidth b, and hyperparameters a_σ, b_σ, α, ρ, a_1, and a_2. In particular, we consider nine scenarios of possible combinations of p and b (Web Table 1) and, for one of this scenarios, we also report results on sensitivities to the prior specifications of the remaining parameters.

To generate a synthetic dataset, we randomly selected 200 studies from the real data analysis (Section 3) whilst keeping the true proportion of sampled emotion studies equal to that of the real data analysis (70% emotion studies, hereafter called "type 1"). We then retained the estimated posterior mean intensity functions at slice $z = -20$ mm as true intensities for the 2D simulated dataset. Given the true intensities, the foci for each study where generated using the spatstat library in R. Web Figure 1 shows the true group average intensity functions and the simulated data points. To simulate the posterior distribution, we run each chain for 50,000 iterations with a burn-in of 20,000, and thinned the chain every 20 iterations to reduce the autocorrelation in the posterior samples. To assess the posterior variability of the intensity functions, Figure 6 shows the posterior histograms of the intensity function evaluated at voxels with highest intensity values for four studies under case scenario 6. The true values fall in the range of posterior samples; and the posterior mean intensities are close to the true values. We examined a variety of different studies

for all scenarios and conclusions were unchanged. Therefore, the method provides a good accuracy in estimating the intensities.

[Figure 6 about here.]

For a numerical comparison across the eleven scenarios, we computed the integrated mean square error (IMSE) between the posterior mean intensity function of for each study and the corresponding true function on axial slice A ($z = -20$ mm),

$$\text{IMSE} = \frac{1}{N} \sum_{i=1}^{N} \int_A \lambda_i(\nu) - \hat{\lambda}_i(\nu) d\nu,$$

where $\hat{\lambda}_i$ is the posterior mean intensity function for study i and λ_i is the true function. Also, we split the data into a train set (50%) and a test set, and computed the AUC to assess the predictive accuracy. Results under the eleven scenarios are reported in Table 1. From these results, it appears that a larger number of bases with relatively narrower kernels has to be preferred to be able to capture of large variety of shapes. In general, however, the values for p and b have to be inferred via sensitivity analysis in that the most appropriate values for these parameters always depend on the application at hand. We observe that results under cases 9-11 are very similar. This implies that the intensity estimates are stable and not sensitive to moderate changes of the model hyperparameters. We report results of a simulation at slice $z = 42$ mm in the Supplementary Materials (Web Appendix B).

[Table 1 about here.]

5   Discussion

The article has proposed a spatial Bayesian latent factor regression model for CBMA data. The basic formulation generalizes the Bayesian latent factor regression model of Montagna et al. (2012), which was developed for the modeling of time-course trajectories, to the analysis of spatial point pattern data for neuroimaging MA. This allows one to include a high-dimensional set of pre-specified basis functions, while allowing automatic shrinkage and effective removal of basis coefficients not needed to characterize any of the study-specific intensity functions. Further, we accommodate joint modeling of a functional predictor with a binary response, the study type, within a framework of

16

scalar-on-image regression. Along the same lines, the proposed framework can be easily modified for joint modeling of data of many different types, e. g., the probit model for a binary outcome can be replaced by an appropriate predictive model for categorical, nominal, or continuous study features. An interesting future direction within our modeling approach is to combine CBMA data with intensity-based meta-analysis (IBMA) data. The volume of literature on IBMA is still limited, though we note that there is a growing interest among researchers in sharing full image data and statistic maps from the studies. The extension to joint modeling of multi-type IBMA and CBMA data will be explored in future research.

Acknowledgements

References

Albert, J. & Chib, S. (1993), 'Bayesian Analysis of Binary and Polychotomous Response Data', *Journal of the American statistical Association* **88**(422), 669–679.
URL: *http://www.tandfonline.com/doi/abs/10.1080/01621459.1993.10476321*

Arminger, G. & Muthén, B. O. (1998), 'A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm', *Psychometrika* **63**(3), 271–300.

Bhattacharya, A. & Dunson, D. B. (2011), 'Sparse Bayesian infinite factor models.', *Biometrika* **98**(2), 291–306.

Brown, G. G., Perthen, J. E., Liu, T. T. & Buxton, R. B. (2007), 'A primer on functional magnetic resonance imaging', *Neuropsychology Review* **17**, 107–125.

Carp, J. (2012), 'The secret lives of experiments: Methods reporting in the fmri literature', *NeuroImage* **63**(1), 289–300.

Cox, D. R. (1955), 'Some statistical methods connected with series of events', *Journal of the Royal Statistical Society Series B* **17**(2), 129–164.

Eickhoff, S. B., Laird, A., Grefkes, C., Wang, L. E., Zilles, K. & Fox, P. T. (2009), 'Coordinate-based activation likelihood estimation meta-analysis of neuroimaging data: a random-effects approach based on empirical estimates of spatial uncertainty', *Human Brain Mapping* **30**, 2907–2926.

Fox, P. T., Lancaster, J. L., Parsons, L. M., Xiong, J. & Zamarripa, F. (1997), 'Functional volumes modeling: Theory and preliminary assessment', *Human Brain Mapping* **5**, 306–311.

Kang, J., Johnson, T. D. & Nichols, T. E. (2014), 'A Bayesian hierarchical spatial point process model for multi-type neuroimaging meta-analysis', *Annals of Applied Statistics* **8**(3), 1800–1824.

Kang, J., Johnson, T. D., Nichols, T. E. & Wager, T. D. (2011), 'Meta-analysis of functional neuroimaging data via Bayesian spatial point processes.', *Journal of the American Statistical Association* **106**(493), 124–134.

Kim, C., Cilles, S. E., Johnson, N. F. & Gold, B. T. (2012), 'Domain general and domain preferential brain regions associated with different types of task switching: a meta-analysis', *Human Brain Mapping* **33**(1), 130–142.

Kober, H., Barrett, L. F., Joseph, J., Bliss-Moreau, E., Lindquist, K. & Wager, T. D. (2008), 'Functional grouping and cortical-subcortical interactions in emotion: A meta- analysis of neuroimaging studies', *NeuroImage* **42**, 998–1031.

Lindquist, K. A., Wager, T. D., Kober, H., Bliss-Moreau, E. & Barrett, L. F. (2012), 'The brain basis of emotion: a meta-analytic review', *Behavioral and Brain Sciences* **35**(3), 121–143.

Møller, J., Syversveen, A. R. & Waagepetersen, R. P. (1998), 'Log gaussian cox processes', *Scandinavian Journal of Statistics* **25**, 451–482.

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

Montagna, S., Tokdar, S. T., Neelon, B. & Dunson, D. B. (2012), 'Bayesian latent factor regression for functional and longitudinal data.', *Biometrics* **68**(4), 1064–73.
URL: *http://www.ncbi.nlm.nih.gov/pubmed/23005895*

Neal, R. M. (2010), MCMC using Hamiltonian dynamics, *in* X.-L. Brooks, Steve; Gelman, Andrew; Jones, Galin; Meng, ed., 'Handbook of Markov Chain Monte Carlo', Boca Raton: Chapman & Hall–CRC Press.

Nee, D. E., Brown, J. W., Askren, M. K., Berman, M. G., Demiralp, E., Krawitz, A. & Jonides, J. (2013), 'A meta-analysis of executive components of working memory', *Cerebral Cortex* **23**(2), 264–282.

Nee, D. E., Wager, T. D. & Jonides, J. (2007), 'Interference resolution: Insights from a meta-analysis of neuroimaging tasks', *Cognitive, Affective, & Behavioral Neuroscience* **7**(1), 1–7.

Nichols, T. E. & Holmes, A. P. (2001), 'Nonparametric permutation tests for functional neuroimaging: a primer with examples', *Human Brain Mapping* **15**, 1–25.

Nielsen, F. A. & Hansen, L. K. (2002), 'Modeling of activation data in the brainmap database: Detection of outliers', *Human Brain Mapping* **15**, 146–156.

Pekar, J. J. (2006), 'A brief introduction to functional mri', *IEEE Engineering in Medicine and Biology Magazine* **25**, 24–26.

Radua, J. & Mataix-Cols, D. (2009), 'Voxel-wise meta-analysis of grey matter changes in obsessive compulsive disorder', *The British Journal of Psychiatry* **195**, 393–402.

Roberts, G. O. & Rosenthal, J. S. (2007), 'Coupling and ergodicity of adaptive Markov Chain Monte Carlo algorithms', *Journal of Applied Probability* **44**(March 2005), 458–475.

Rottschy, C., Langner, R., Dogan, I., Reetz, K., Laird, a. R., Schulz, J. B., Fox, P. T. & Eickhoff, S. B. (2012), 'Modelling neural correlates of working memory: a coordinate-based meta-analysis', *NeuroImage* **60**, 830–846.

Rue, H. & Held, L. (2005), *Gaussian Markov Random Fields*, Chapman and Hall/CRC.

Samartsidis, P., Montagna, S., Johnson, T. D. & Nichols, T. E. (2016), The coordinate-based meta-analysis of fmri data, Technical report, University of Warwick.

Taylor, B. M. & Diggle, P. J. (2014), 'Inla or mcmc? a tutorial and comparative evaluation for spatial prediction in log-gaussian cox processes', *Journal of Statistical Computation and Simulation* **84**, 2266–2284.

Thirion, J., Pinel, P., Meriaux, S., Roche, A., Dehaene, S. & Poline, J.-B. (2007), 'Analysis of a large fmri cohort: Statistical and methodological issues for group analyses', *NeuroImage* **35**, 105–120.

Turkeltaub, P. E., Eden, G. F., Jones, K. M. & Zeffiro, T. A. (2002), 'Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation', *NeuroImage* **16**, 765–780.

Wager, T. D., Jonides, J. & Reading, S. (2004), 'Neuroimaging studies of shifting attention: a meta-analysis', *NeuroImage* **22**, 1679–1693.

Wager, T. D., Kang, J., Johnson, T. D., Nichols, T. E., Satpute, A. B. & Barrett, L. F. (2015), 'A bayesian model of category-specific emotional brain responses', *PLOS Computational Biology* **11**(4), e1004066.

Wager, T. D., Lindquist, M. & Kaplan, L. (2007), 'Meta-analysis of functional neuroimaging data: current and future directions', *Social Cognitive and Affective Neuroscience* **2**(2), 150–158.

Wager, T. D. & Smith, E. E. (2003), 'Neuroimaging studies of working memory', *Cognitive, Affective, and Behavioral Neuroscience* **3**(4), 255274.

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

Wood, A. & Chan, G. (1994), 'Simulation of stationary gaussian processes in $[0,1]^d$', *Journal of Computational and Graphical Statistics* **3**, 409–432.

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J. & Evans, A. C. (1996), 'A unified statistical approach for determining significant signals in images of cerebral activation', *Human Brain Mapping* **4**, 58–73.

Yarkoni, T., Poldrack, R. a., Nichols, T. E., Van Essen, D. C. & Wager, T. D. (2011), 'Large-scale automated synthesis of human functional neuroimaging data.', *Nature methods* **8**(8), 665–70.

Yeo, T. B. T., Krienen, F. M., Eickhoff, S. B., Yaakub, S. N., Fox, P. T., Buckner, R. L., Asplund, C. L. & Chee, M. W. L. (2015), 'Functional specialization and flexibility in human association cortex', *Cerebral Cortex* **25**(10), 3654–3672.

Yue, Y. U., Lindquist, M. A. & Loh, J. M. (2012), 'Meta-analysis of functional neuroimaging data using bayesian nonparametric binary regression', *Annals of Applied Statistics* **6**, 697–718.

Table 1: Simulation study results. Comparison of the IMSE and AUC summary measures for our model under the eleven scenarios in Web Table 1. We report the IMSE relative to the value obtained for case scenario 9 (rIMSE).

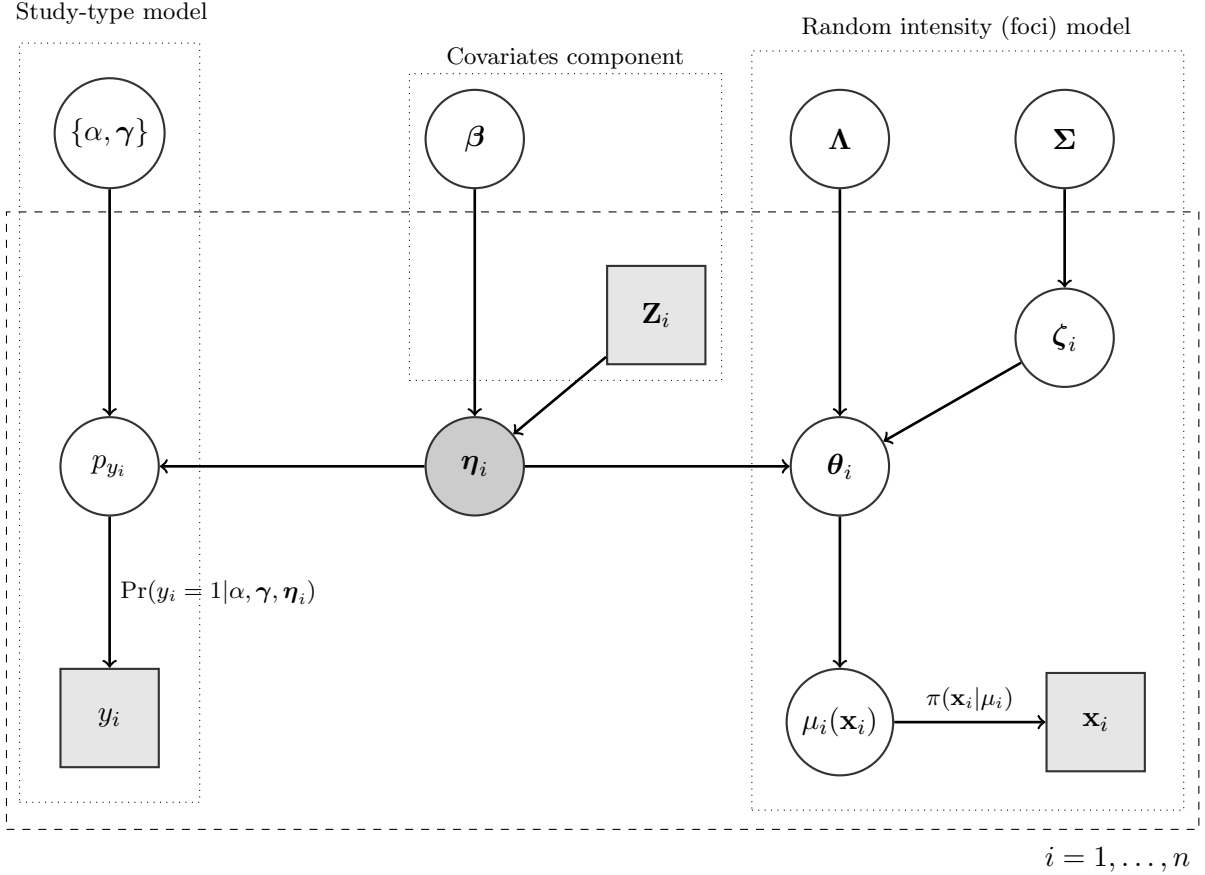| Scenarios	rIMSE	AUC
1	1.181	0.57
2	1.186	0.58
3	1.312	0.49
4	1.147	0.62
5	1.013	0.60
6	1.004	0.65
7	1.233	0.62
8	1.122	0.65
9	**1**	**0.65**
10	1.011	0.65
11	1.061	0.65

Figure 1: Graphical representation of the probabilistic mechanism generating data $\{\mathbf{x}_i, y_i\}$, $i = 1, \ldots, n$, under the spatial Bayesian latent factor model. Shaded squares represent observed quantities and circles represent unknowns. The circle denoting the vector of latent factors is darkened. Note that the study type, y_i, is conditionally independent of all other nodes in the "random intensity model" given the latent variables $\boldsymbol{\eta}_i$.
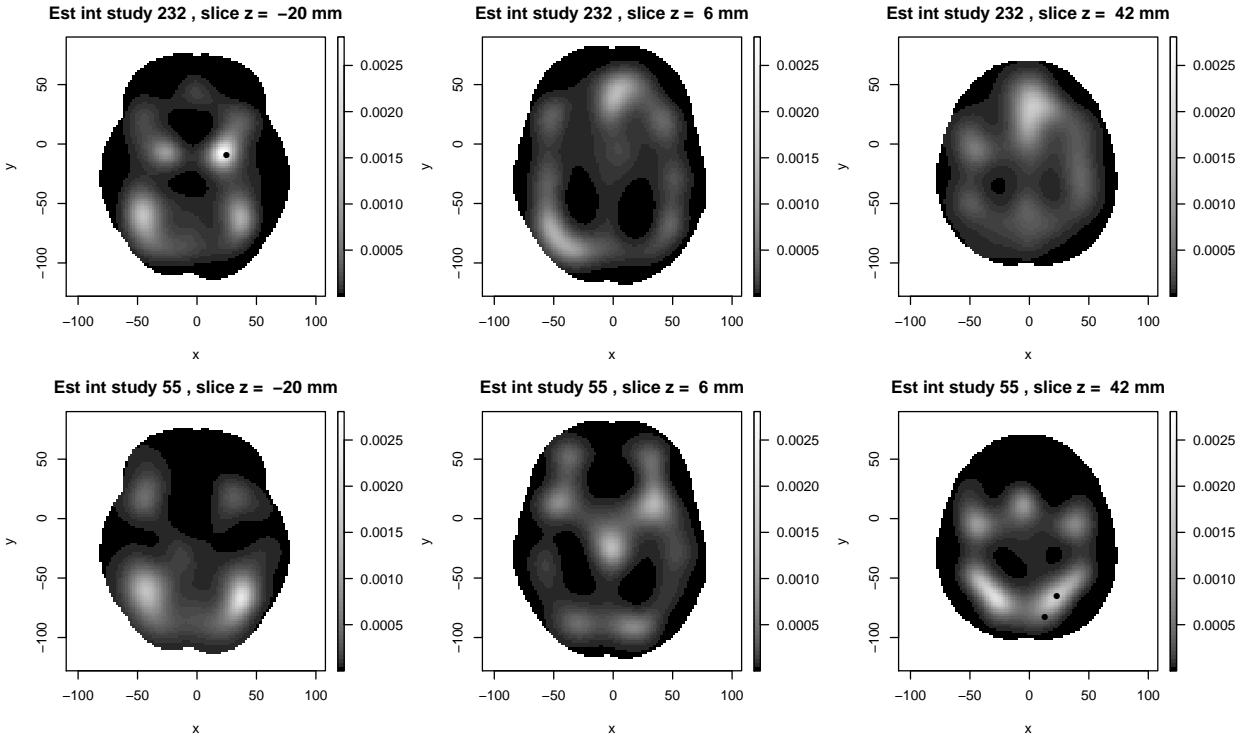
Figure 2: Posterior mean intensity estimates for two randomly chosen studies. Top row: an emotion study; bottom row: an executive control study. Here we only show three axial slices (columns) of the fully 3D results. Black points denote reported foci at the corresponding axial slices.
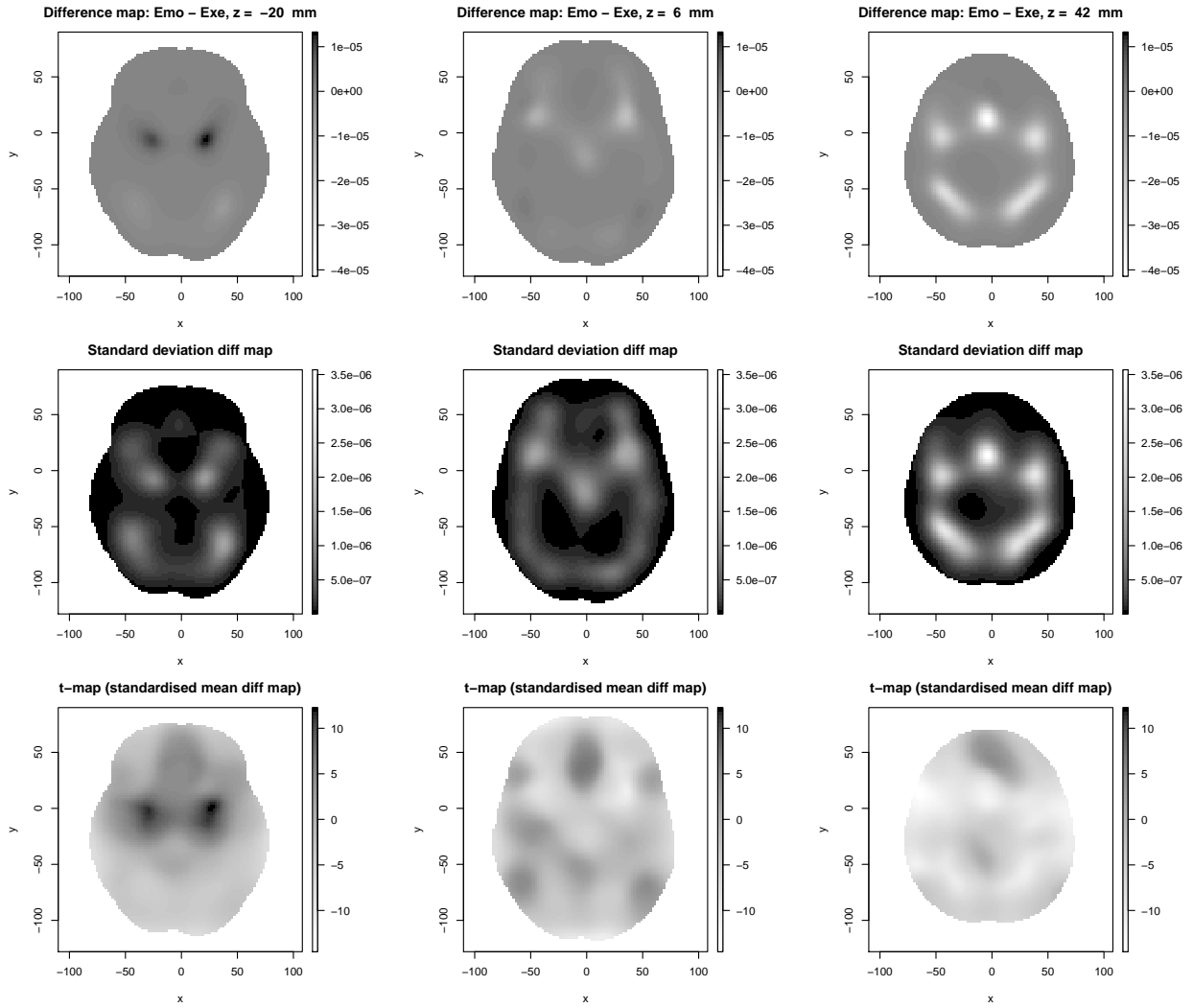
Figure 3: Top row: posterior mean of difference map for the emotion studies posterior group intensity vs. executive control studies posterior group intensity; middle row: estimated standard deviation of difference map; bottom row: standardised mean difference map. Here we only show three axial slices (columns) of the fully 3D results. The grey color scale for the middle row has been reversed to improve visibility of the results.
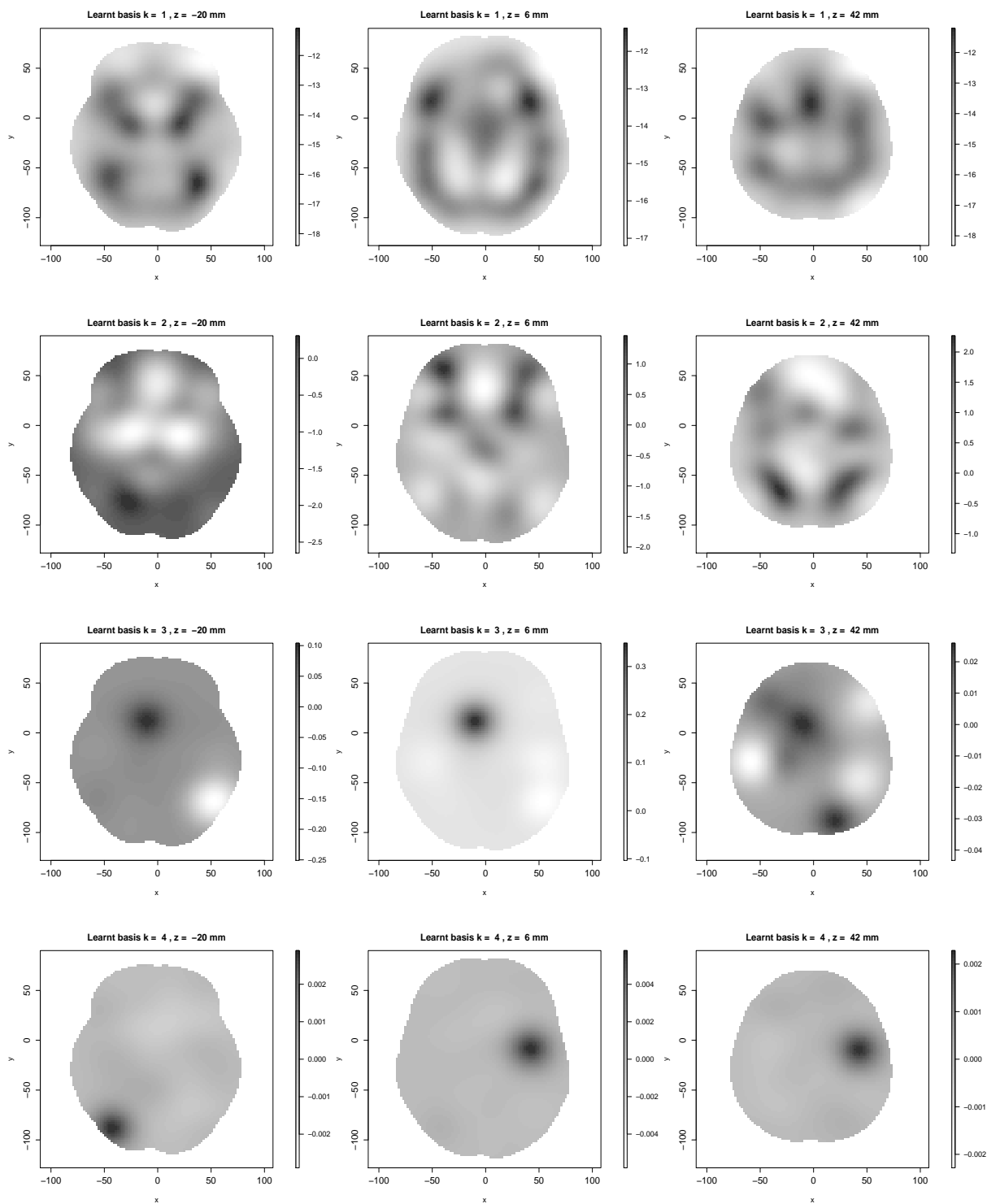
Figure 4: Learnt dictionary elements $\{\phi_l\}_{l=1}^4$ at three axial slices (columns). The estimated posterior mean number of factors is $k = 4$.
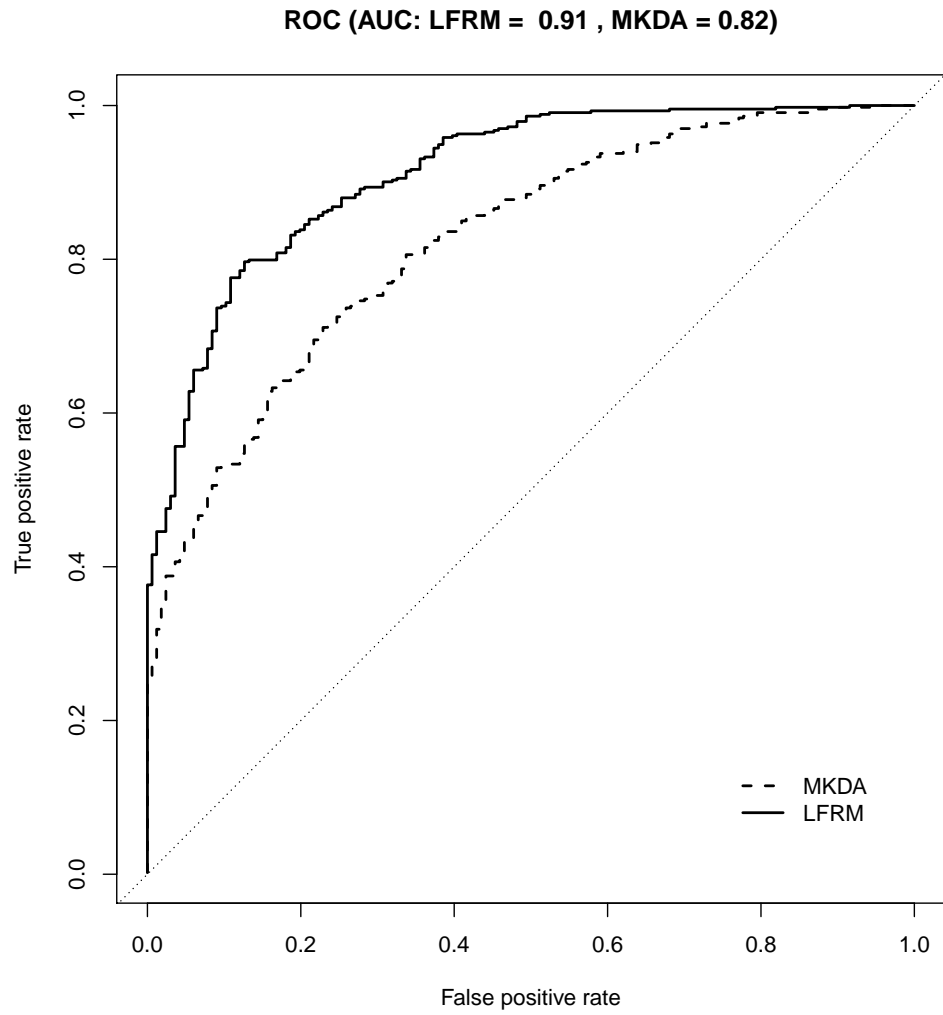
Figure 5: ROC curve for prediction of study type for studies in the test set. The AUC corresponds to the area under the curve.
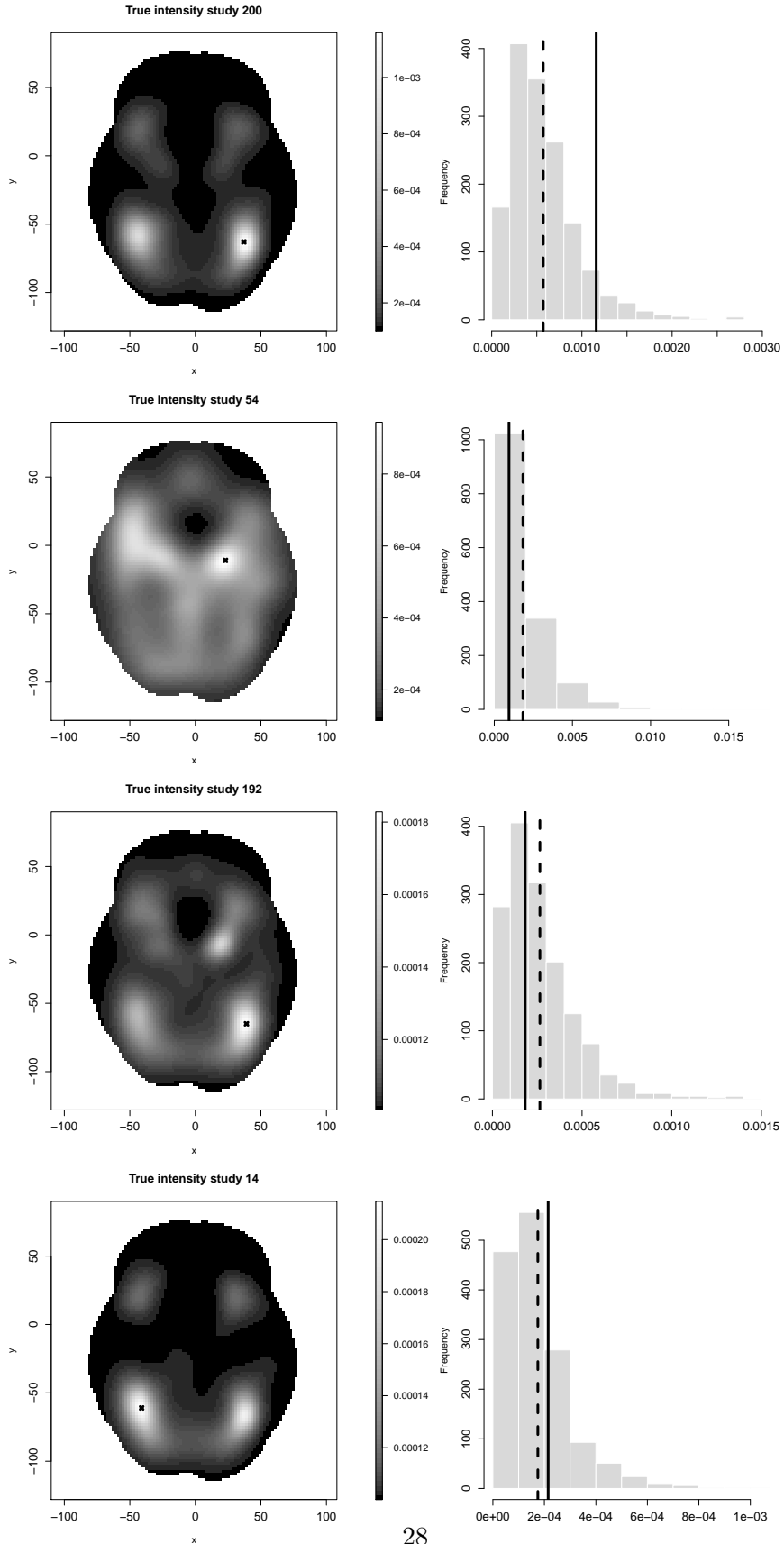
Figure 6: True intensity function for four randomly chosen studies in the simulated dataset. Histograms show the posterior distribution of the intensity function evaluated at voxels of highest intensity values (black "x"). The dashed line is the marginal posterior mean and the solid line is the true intensity.

Supplementary Materials for "Spatial Bayesian Latent Factor Regression Modeling of Coordinate-based Meta-analysis Data" by Silvia Montagna, Tor Wager, Lisa Feldman Barrett, Timothy D. Johnson, and Thomas E. Nichols

SILVIA MONTAGNA

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

TOR WAGER

Department of Psychology and Neuroscience, University of Colorado at Boulder, Boulder, CO 80309, U.S.A.

LISA FELDMAN BARRETT

Department of Psychology, Northeastern University, Boston, MA 02115, U.S.A.

TIMOTHY D. JOHNSON

Biostatistics Department, University of Michigan, Ann Arbor, MI 48109, U.S.A.

THOMAS E. NICHOLS

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

Web Appendix A

Posterior computation

We provide a description of the MCMC algorithm used to update from the posterior distribution of the model parameters in Section 2.1 for the LFRM on the random Poisson intensities. The sampler cycles through the following steps:

1. *Update of $\boldsymbol{\theta}_i$*: Conditioning on all other model parameters, the log-posterior is given by

$$\log \pi(\boldsymbol{\theta}_i \mid -) \propto -\int_{\mathcal{B}} \exp\{\mathbf{b}(\mathbf{s})^\top \boldsymbol{\theta}_i\} ds + \sum_{j=1}^{n_i} \mathbf{b}(\mathbf{x}_{ij})^\top \boldsymbol{\theta}_i - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i),$$

for $i = 1, \ldots, n$. We recall that $\{\mathbf{x}_{ij}\}_{j=1}^{n_i}$ denotes the set of foci reported by study i. After discretising to a grid $B \subseteq \mathcal{B}$, the log posterior becomes

$$\log \pi(\boldsymbol{\theta}_i \mid -) \propto -V \sum_{l \in B} \exp\{\mathbf{b}(\mathbf{l})^\top \boldsymbol{\theta}_i\} + \sum_{j=1}^{n_i} \mathbf{b}(\mathbf{x}_{ij})^\top \boldsymbol{\theta}_i - \frac{1}{2}(\boldsymbol{\theta}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)^\top \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i), \quad (1)$$

where V is the volume of each voxel. We resort to HMC [2] to sample from (1) by defining the potential energy function as

$$U(\boldsymbol{\theta}_i) = -\log \pi(\boldsymbol{\theta}_i \mid -),$$

1

and its derivative with respect to $\boldsymbol{\theta}_i$ corresponds to

$$\frac{\partial U}{\partial \boldsymbol{\theta}_i} = V \sum_{l \in B} \mathbf{b}(\mathbf{l})^\top \exp\{\mathbf{b}(\mathbf{l})^\top \boldsymbol{\theta}_i\} - \sum_{j=1}^{n_i} \mathbf{b}(\mathbf{x}_{ij})^\top + (\boldsymbol{\theta}_i - \boldsymbol{\Lambda}\boldsymbol{\eta}_i)^\top \boldsymbol{\Sigma}^{-1}.$$

The kinetic energy, $K(\tilde{\boldsymbol{\theta}}_i)$, is assumed to have form $K(\tilde{\boldsymbol{\theta}}_i) = \sum_{m=1}^p \frac{\tilde{\theta}_{ij}^2}{2}$. A detailed presentation of Hamiltonian dynamics is beyond the scope of this paper. We defer to [2] for a description of the leapfrog method and further details

2. *Update of* $\boldsymbol{\Lambda}$: Sample $\lambda_{jh}, \delta_1, \delta_h, \iota_{jh}$ from the following posteriors:

   (a) Denote the jth row of $\boldsymbol{\Lambda}_{k^*}$ (the loading matrix $\boldsymbol{\Lambda}$ truncated to $k^* \ll p$) by $\boldsymbol{\lambda}_j$; the $\boldsymbol{\lambda}_j$'s have conditionally independent conjugate posteriors given by

   $$\pi(\boldsymbol{\lambda}_j \mid -) \sim N_{k^*}((\mathbf{D}_j^{-1} + \sigma_j^{-2}\boldsymbol{\eta}^\top\boldsymbol{\eta})^{-1}\boldsymbol{\eta}^\top \sigma_j^{-2}\boldsymbol{\theta}^{(j)}, (\mathbf{D}_j^{-1} + \sigma_j^{-2}\boldsymbol{\eta}^\top\boldsymbol{\eta})^{-1})$$

   with $\mathbf{D}_j^{-1} = \mathrm{diag}(\iota_{j1}\tau_1, \dots, \iota_{jk^*}\tau_{k^*})$, $\boldsymbol{\eta}^\top = [\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_{k^*}]$ and $\boldsymbol{\theta}^{(j)} = (\theta_{j1}, \dots, \theta_{jn})$, for $j = 1, \dots, p$

   (b) Sample ι_{jh} from

   $$\pi(\iota_{jh} \mid -) \sim \mathrm{Gamma}\left(\frac{\rho+1}{2}, \frac{\rho}{2} + \frac{\tau_h \lambda_{jh}^2}{2}\right)$$

   (c) Sample δ_1 from

   $$\pi(\delta_1 \mid -) \sim \mathrm{Gamma}\left(a_1 + \frac{pk^*}{2}, 1 + \frac{1}{2}\sum_{l=1}^{k^*} \tau_l^{(1)} \sum_{j=1}^p \iota_{jl}\lambda_{jl}^2\right),$$

   and for $h \geqslant 2$, sample δ_h from

   $$\pi(\delta_h \mid -) \sim \mathrm{Gamma}\left(a_2 + \frac{p}{2}(k^* - h + 1), 1 + \frac{1}{2}\sum_{l=h}^{k^*} \tau_l^{(h)} \sum_{j=1}^p \iota_{jl}\lambda_{jl}^2\right),$$

   where $\tau_l^{(h)} = \prod_{t=1, t \neq h}^l \delta_t$ for $h = 1, \dots, k^*$

   The sampling begins with a very conservative choice of k^*, which is then automatically selected within the adaptive Gibbs sampler as described in [1]

3. *Update of* $\boldsymbol{\eta}_i$: Sample $\boldsymbol{\eta}_i$ from the full conditional posterior

   $$\pi(\boldsymbol{\eta}_i \mid -) \sim N\left(\left(\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} + \boldsymbol{I}_k\right)^{-1}(\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_i + \boldsymbol{\beta}^\top\mathbf{Z}_i), \left(\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} + \boldsymbol{I}_k\right)^{-1}\right)$$

   for $i = 1, \dots, n$. We recall that $\mathbf{Z}_i$ is the $r \times 1$ vector of covariates for study i

4. *Update of* σ_j^2: Denote with $\sigma_j^{-2}, j = 1, \dots, p$, the diagonal elements of $\boldsymbol{\Sigma}^{-1}$. Assume $\sigma_j^{-2} \sim \mathrm{Gamma}(a_\sigma, b_\sigma)$. Sample σ_j^{-2} from conditionally independent posteriors

   $$\pi(\sigma_j^{-2} \mid -) \sim \mathrm{Gamma}\left(a_\sigma + \frac{n}{2}, b_\sigma + \frac{\sum_{i=1}^n (\theta_{ij} - \boldsymbol{\lambda}_j\boldsymbol{\eta}_i)^2}{2}\right)$$

   where $\boldsymbol{\lambda}_j$ corresponds to the jth row of $\boldsymbol{\Lambda}$

2

5. *Update of $\boldsymbol{\beta}$:* A Cauchy prior is induced on the columns of the $r \times k$ matrix of coefficients as follows

$$\boldsymbol{\beta}_l \sim N(\mathbf{0}, \mathrm{Diag}(w_{lj}^{-1})), \quad \text{with} \quad w_{lj} \sim \mathrm{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right)$$

for $l = 1, \ldots, k$ and $j = 1, \ldots r$. The update proceeds by sampling

(a) ω_{lj} from its full conditional posterior

$$\pi(\omega_{lj} \mid -) \sim \mathrm{Gamma}\left(1, \frac{1}{2}\left(1 + \beta_{lj}^2\right)\right)$$

(b) the lth column of $\boldsymbol{\beta}$ from its full conditional posterior

$$\pi(\boldsymbol{\beta}_l \mid -) \sim N\left(\left(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top + \mathbf{E}_l^{-1}\right)^{-1}\tilde{\mathbf{Z}}\boldsymbol{\eta}_{\cdot l}^\top, \left(\tilde{\mathbf{Z}}\tilde{\mathbf{Z}}^\top + \mathbf{E}_l^{-1}\right)^{-1}\right),$$

where $\boldsymbol{\eta}_{\cdot l}^\top \sim N(\tilde{\mathbf{Z}}^\top\boldsymbol{\beta}_l, \boldsymbol{I}_n)$ denotes the lth column of the $n \times k$ transpose of the matrix of latent factors $\boldsymbol{\eta}$, $\boldsymbol{I}_n$ is the $n \times n$ identity matrix, and $\tilde{\mathbf{Z}}^\top$ denotes the transpose of the $r \times n$ matrix of predictors $\tilde{\mathbf{Z}}$. Each row i of $\tilde{\mathbf{Z}}^\top$ corresponds to the vector of covariates for study i, for $i = 1, \ldots, n$. Matrix $\mathbf{E}_l$ corresponds to $\mathbf{E}_l = \mathrm{Diag}(\omega_{lj}^{-1})$, for $l = 1, \ldots, k$ and $j = 1, \ldots, r$

The probit extension described in Section 2.2 involves a straightforward modification of the MCMC algorithm described above, which now includes additional steps to sample from the full conditional posterior distributions of the new model parameters. To account for the study-type component model, the update of the latent factors is modified as follows

(3b) *Update of $\boldsymbol{\eta}_i$:* Sample $\boldsymbol{\eta}_i$ from a multivariate Gaussian full conditional posterior distribution with mean

$$\mathbb{E}(\boldsymbol{\eta}_i|-) = \left(\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} + \boldsymbol{I}_k + \boldsymbol{\gamma}\boldsymbol{\gamma}^\top\right)^{-1}(\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\theta}_i + \boldsymbol{\beta}^\top\mathbf{Z}_i + \boldsymbol{\gamma}(W_i - \alpha)),$$

and covariance

$$\mathbb{C}\mathrm{ov}(\boldsymbol{\eta}_i|-) = \left(\boldsymbol{\Lambda}^\top\boldsymbol{\Sigma}^{-1}\boldsymbol{\Lambda} + \boldsymbol{I}_k + \boldsymbol{\gamma}\boldsymbol{\gamma}^\top\right)^{-1},$$

for $i = 1, \ldots, n$

Further, we have the following three additional steps:

(6) *Update of $\boldsymbol{\gamma}$:* Sample $\boldsymbol{\gamma}$ from the following full conditional posterior distribution

$$\pi(\boldsymbol{\gamma}|-) \sim N\left((\boldsymbol{\Sigma}_\gamma^{-1} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)^{-1}(\boldsymbol{\Sigma}_\gamma^{-1}\boldsymbol{\mu}_\gamma + \boldsymbol{\eta}(\mathbf{W} - \boldsymbol{\alpha})), (\boldsymbol{\Sigma}_\gamma^{-1} + \boldsymbol{\eta}\boldsymbol{\eta}^\top)^{-1}\right),$$

where $\mathbf{W} = [W_1, \ldots, W_n]^\top$ and $\boldsymbol{\alpha}$ is the $n \times 1$ vector of n replicates of α, $\boldsymbol{\alpha} = [\alpha, \ldots, \alpha]_{n \times 1}^\top$

(7) *Update of α:* Sample α from the following full conditional posterior distribution

$$\pi(\alpha|-) \sim N\left(\frac{v_\alpha}{1 + v_\alpha n} \times \left(\frac{\mu_\alpha}{v_\alpha} + \sum_{i=1}^n \left(W_i - \boldsymbol{\eta}_i^\top\boldsymbol{\gamma}\right)\right), \frac{v_\alpha}{1 + v_\alpha n}\right)$$

(8) *Update of W_i*: Sample W_i from the following full conditional posterior distribution

$$\pi(W_i|-) = \begin{cases} TN_{[0,\infty)}(\alpha + \boldsymbol{\gamma}^\top \boldsymbol{\eta}_i, 1) & \text{if } y_i = 1 \\ TN_{(-\infty,0)}(\alpha + \boldsymbol{\gamma}^\top \boldsymbol{\eta}_i, 1) & \text{if } y_i = 0, \end{cases}$$

for $i = 1, \ldots, n$, where $TN_{[a,b]}$ denotes the truncated normal distribution on the interval $[a, b]$

Web Appendix B

Simulation study (cont.'d Section 4)

Web Table 1 and Web Figure 1 refer to the simulation study presented in Section 4 of the paper.

Web Table 1: Prior specifications of eleven scenarios for sensitivity analysis.

| Scenarios	p	b	a_σ	b_σ	m_α	a_1	a_2	ρ
1	25	1/800	1	0.3	$\Phi^{-1}(0.5)$	2.1	3.1	3
2	25	1/512			$\vdots$			
3	25	1/128			$\vdots$			
4	48	1/800			$\vdots$			
5	48	1/512			$\vdots$			
6	48	1/128			$\vdots$			
7	90	1/800			$\vdots$			
8	90	1/512			$\vdots$			
9	90	1/128	1	0.3	$\Phi^{-1}(0.5)$	2.1	3.1	3
10	90	1/128	1	0.25	$\Phi^{-1}(0.62)$	3.1	2.1	2
11	90	1/128	2	2	$\Phi^{-1}(0.62)$	3.1	2.1	2

Sensitivity analysis

We report results on a second simulation study for sensitivity analysis. Data was generated as described in Section 4, but we now retained the estimated marginal posterior mean intensity functions at slice $z = 42$ mm as true intensities for the 2D dataset. Web Figure 2 shows the true group average intensity functions and simulated data points.

To simulate the posterior distribution, we run the chain for 50,000 iterations with a burn-in of 20,000, and thinned the chain every 20 iterations to reduce the autocorrelation in the posterior samples. To assess the posterior variability of the intensity functions, Web Figure 3 shows the posterior histograms of the intensity function evaluated at voxels with highest intensity values for four studies. The true values fall in the range of posterior samples; and the posterior mean intensities are close to the true values. We examined a variety of other different studies and
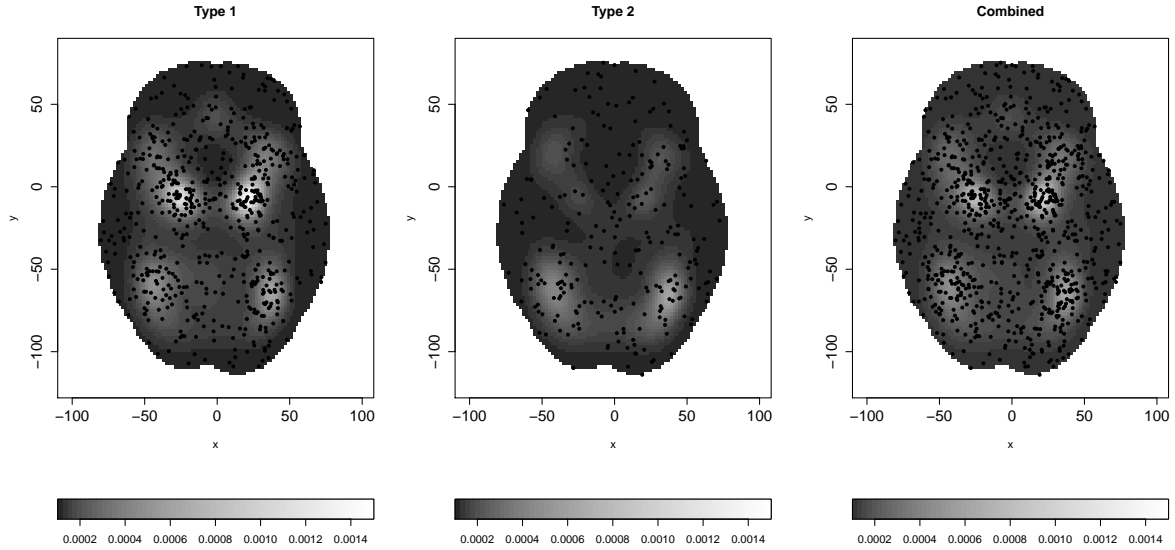
Web Figure 1: True group average intensity functions with data points for the simulation study in Section 4 of the paper.
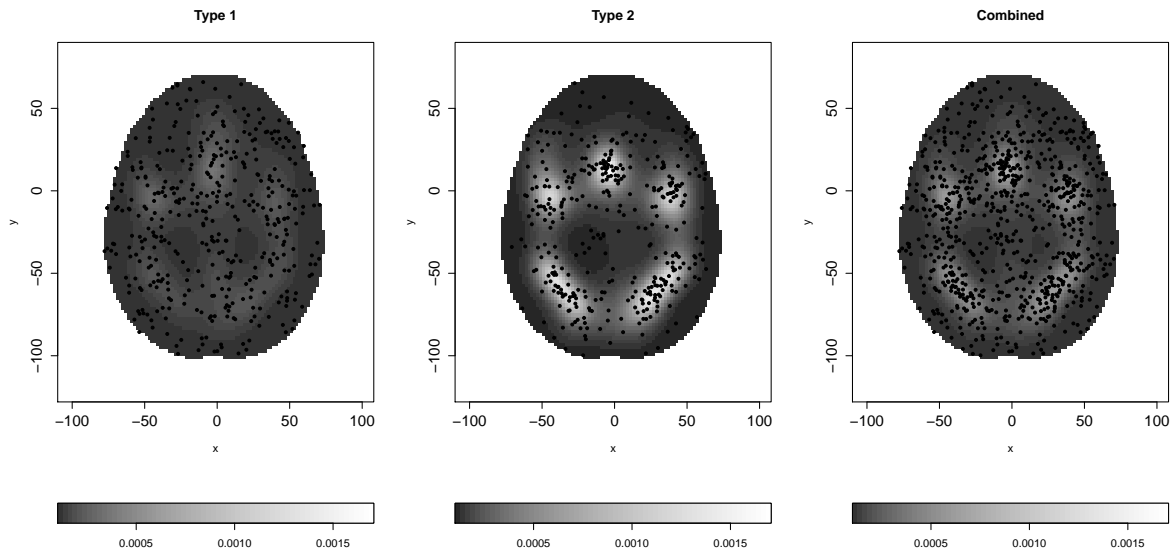
Web Figure 2: True group average intensity functions with simulated data points at slice $z = 42$ mm.

conclusions were unchanged. Therefore, the method provides a good accuracy in estimating the intensities.

Web Table 2 shows a comparison of the relative IMSE and AUC summary measures for our model under nine scenarios of possible combinations of p and b. Again, the studies were split into a train set (50%) and a test set to assess the predictive accuracy. As we observed in Section 4, large values of p with relatively narrow kernels are to be preferred to obtain a reduction in IMSE to the truth. However, the AUC does not vary significantly across the nine

5
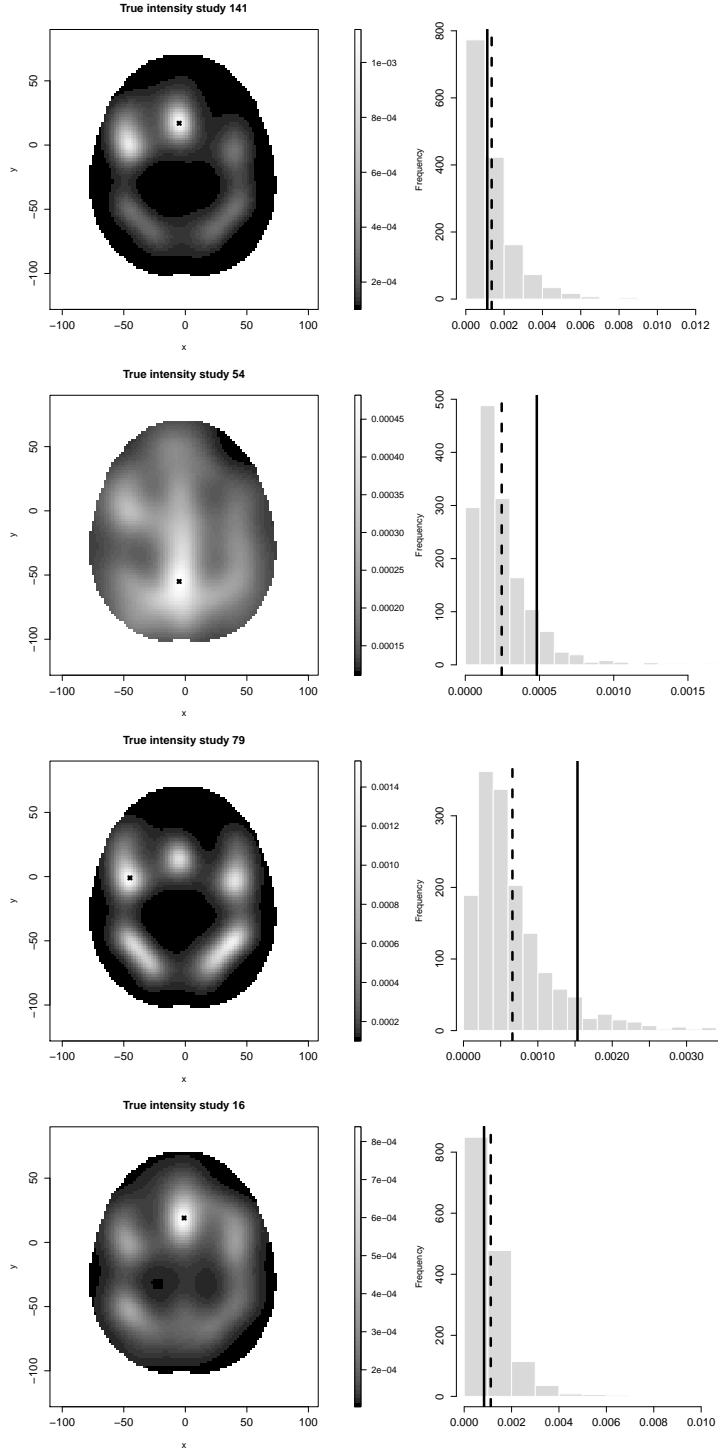
Web Figure 3: True intensity function for four randomly chosen studies in the simulated dataset at $z = 42$ mm. Histograms show the posterior distribution of the intensity function evaluated at voxels of highest intensity values (black "x"). The dashed line is the marginal posterior mean and the solid line is the true intensity.

Web Table 2: Simulation study results ($z = 42$ mm). Comparison of the IMSE and AUC summary measures for our model under nine scenarios of possible combinations of p and b. We report the IMSE relative to the value obtained for case scenario 9 (rIMSE).

Scenarios	p	b	rIMSE	AUC
1	21	1/800	1.230	0.74
2	21	1/512	1.272	0.72
3	21	1/128	1.413	0.75
4	45	1/800	1.115	0.73
5	45	1/512	1.034	0.74
6	45	1/128	1.157	0.73
7	86	1/800	1.264	0.72
8	86	1/512	1.146	0.74
9	86	1/128	**1**	**0.74**

scenarios, and it is higher than the AUC obtained for the simulations at slice $z = -20$ mm. A possible is explanation is that the foci are more evenly spread between the two study types at this slice (Web Figure 2), whereas the type 1 studies in Section 4 have more foci as compared to type 2 (Web Figure 1), and therefore the predictive classification is more likely to be biased toward type 1.

References

[1] Anirban Bhattacharya and David B Dunson. Sparse Bayesian infinite factor models. *Biometrika*, 98(2):291–306, June 2011.

[2] Radford M. Neal. MCMC using Hamiltonian dynamics. In Xiao-Li Brooks, Steve; Gelman, Andrew; Jones, Galin; Meng, editor, *Handbook of Markov Chain Monte Carlo*. Boca Raton: Chapman & Hall–CRC Press, 2010.