

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

Silvia Montagna^{*1}, Tor Wager², Lisa Feldman-Barrett³, Timothy D. Johnson⁴ and Thomas
E. Nichols⁵

¹School of Mathematics, Statistics and Actuarial Science, University of Kent, Canterbury
CT2 7FS, U.K.

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

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

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

⁵Department of Statistics, University of Warwick, Coventry CV4 7AL, U.K.

March 23, 2018

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

^{*}silvia.montagna@unito.it

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 neuroimaging meta-analysis datasets.

Keywords: Bayesian modeling; Factor analysis; Functional principal component analysis; Meta-analysis; Spatial point pattern data; Reverse inference.

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 typically determined via random field theory [Worsley et al., 1996] or permutation methods [Nichols and 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; crucially 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 of neuroimaging studies. By combining the results of independently conducted studies, meta-analysis 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 (a.k.a. forward inference), 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. 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, a single region may be activated by a range of different tasks [Yeo et al., 2015].

Given that published fMRI studies rarely share the statistic images or raw data, meta-analysis 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 meta-analysis based on foci as Coordinate-Based Meta-Analysis (CBMA). Several approaches to CBMA can be found in the literature. See, for example, Turkeltaub et al. [2002], Wager et al. [2007], Kober et al. [2008], Eickhoff et al. [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 (refer to Samartsidis et al. [2016] for an extensive review).

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 (meta-analysis 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 meta-analysis maps. These approaches, however, have some serious limitations. In particular, they are based on a MUA that lacks an explicit spatial approach to the modeling of the foci. As opposed to generative, multivariate (spatial stochastic) models, kernel-based methods do not provide an accurate representation of the true data generating mechanism (non-generative methods) as they do not jointly characterize randomness of the number and locations of activations within each study. Further, these methods do not provide any measure of uncertainty associated with the effect estimate, and conclusions could be misleading. For example, Samartsidis et al. [2016] show that power properties of ALE do not degrade with the inclusion of poor quality studies. Finally, kernel-based methods require ad-hoc spatial kernel parameters (full width half max, FWHM) that need to be pre-specified, and a poor choice for the kernel size could potentially affect the results. In particular, Tench et al. [2014] show that fixing the kernel-size can result in increased false positives as the number of studies in the meta-analysis increases. To overcome this limitation, the authors redefine the FWHM parameter as function of the number of studies in the analysis and provide a method to estimate it.

Recently, model-based approaches have been proposed to overcome some of the limitations of kernel-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 meta-analysis 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 meta-analysis to simultaneously consider several types of tasks. To address this limitation, Kang et al. [2014] generalize the Poisson/gamma random field (PGRF) model of Wolpert and Ickstadt [1998] to a Bayesian hierarchical PGRM fit to multi-type meta-analyses. In particular, the authors regard the foci from each type of study as a realization of a Poisson point process driven by a type-specific random intensity function, which is modeled as a kernel convolution of a type-specific gamma random field. These type-specific gamma random fields are modeled as a realization of a common gamma random field shared by all types (hence the hierarchy), thus introducing dependence between types. Also, the authors propose a model-based classifier to perform reverse inference. While the hierarchical PGRF is a flexible non-parametric model, it relies on highly advanced mathematical and statistical modeling that could be less interpretable and more difficult to communicate to a less-technical audience. Also, this model is difficult to re-implement if software is not made available. Finally, the model does not accommodate covariate information, though an extension to meta-regression is possible.

In this paper, we propose a Bayesian hierarchical model that extends the Bayesian latent factor regression model for longitudinal data of Montagna et al. [2012] to the analysis of CBMA data. In particular, we model the foci from each study as a “doubly stochastic” Poisson process [Cox, 1955], where the study-specific log intensity function is characterised as a linear combination of a 3-dimensional basis set. We induce sparsity on the basis function coefficients via a latent factor model, and information on covariates is incorporated through a simple linear regression model on the latent factors. Further, the latent factors are used as a vehicle to link the intensity functions to a study-type as part of a scalar-on-image regression. 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. We illustrate our approach on a functional neuroimaging meta-analysis of emotions first reported in Kober et al. [2008]. We focus on a subset of the original dataset that consists of 187 studies on five emotions

(sad, happy, anger, fear, and disgust) reporting a total of 984 foci. The goal is to find consistent regions of activation across the different studies and types of emotions.

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 functional principal component analysis. In Section 3, we apply our model to the meta-analysis dataset of emotion studies, and compare our results with MKDA. We conclude the manuscript with a final discussion of our model (Section 4).

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 methodology by presenting an analogy with functional principal component analysis (fPCA). Here the word “functional” stems from the application of PCA to random functions such as curves or any data object varying over a continuum, distinct from “functional” in fMRI.

2.1 The model

Consider independent spatial point patterns arising from n studies, $\mathbf{x}_1, \dots, \mathbf{x}_n$. 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

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

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 μ_1, \dots, μ_n , we follow Montagna et al. [2012]. Specifically, we write $\log \mu_i$ in terms of a collection of basis functions

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

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, B-splines or Gaussian kernels can be used to model smooth μ_i intensities. Hereafter, we use 3D isotropic Gaussian kernels

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

with kernel locations $\{\boldsymbol{\psi}_m\}_{m=2}^p$ and bandwidth b to be specified according to prior knowledge (refer to Web Appendix A for a discussion). 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. We remark that we also implemented our model using B-splines and we did not find

significant differences in the results reported hereafter.

Representation (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 and Chan, 1994, Rue and Held, 2005]. 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 basis representation, all variation between the study-specific intensities are reflected through the variation in the vectors $\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_n$. However, the basis function approach fails 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 other variables under study. In our construction, the μ_i 's are represented by the long vector of coefficients $(\theta_{i1}, \dots, \theta_{ip})$. 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 and Muthén, 1998] on the basis coefficients

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

where $\boldsymbol{\theta}_i = [\theta_{i1}, \dots, \theta_{ip}]^\top$, $\mathbf{\Lambda}$ is a $p \times k$ factor loading matrix with $k \ll p$, $\boldsymbol{\eta}_i = (\eta_{i1}, \dots, \eta_{1k})^\top$ is a vector of latent factors for study i , and $\boldsymbol{\zeta}_i = (\zeta_{i1}, \dots, \zeta_{ip})^\top$ 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 $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_p^2)$. Vectors $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n$ 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

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

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

$$\text{Mean}[\log \mu_i(\boldsymbol{\nu})] = \sum_{l=1}^k \boldsymbol{\beta}_l^\top \mathbf{Z}_i \phi_l(\boldsymbol{\nu}) = \sum_{j=1}^r Z_{ij} \times \left[\sum_{l=1}^k \beta_{jl} \phi_l(\boldsymbol{\nu}) \right] \quad (6)$$

and common covariance function

$$\text{Cov}\{\log \mu_i(\boldsymbol{\nu}), \log \mu_i(\boldsymbol{\nu}')\} = \sum_{l=1}^k \phi_l(\boldsymbol{\nu}) \phi_l(\boldsymbol{\nu}') + \sum_{m=1}^p \sigma_m^2 b_m(\boldsymbol{\nu}) b_m(\boldsymbol{\nu}'), \quad (7)$$

where $\phi_l(\boldsymbol{\nu}) = \sum_{m=1}^p \lambda_{lm} b_m(\boldsymbol{\nu})$, and $\boldsymbol{\beta}_l$ is the l th column of $\boldsymbol{\beta}$. Equation (6) shows how the matrix of covariate coefficients $\boldsymbol{\beta}$ impacts on the expected log intensity function. In particular, $\sum_{l=1}^k \beta_{jl} \phi_l(\boldsymbol{\nu})$ quantifies the expected difference in the mean log intensity function at voxel $\boldsymbol{\nu}$ for a one-unit change in the value of covariate j , with all other quantities being equal. Maps $\sum_{l=1}^k \beta_{jl} \phi_l(\cdot)$ will be shown for the real data application in Section 3. The use of Gaussian-shaped basis functions (centred densely in the $\boldsymbol{\nu}$ -space) guarantees that the covariance function in (7) corresponds to the stationary squared-exponential covariance function [Mackay, 1998, Rasmussen and Williams, 2005]. If a non-stationary covariance is warranted, multiresolution (wavelet) basis functions could be alternatively considered.

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 $y_i \in \{1, \dots, T\}$. 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 focusing on our motivating application, a meta-analysis of emotions first reported in Kober et al. [2008]. We use a subset of the data and focus on five emotions: sad, happy, anger, fear, and disgust. To predict the emotion elicited in a newly presented study, we need to build a predictive model for the study type. In a recent contribution, Johndrow et al. [2013] developed the diagonal orthant multinomial (DO) models, a new class of models to the Bayesian classification of unordered categorical response data. DO models circumvent the traditional limitations faced by multinomial logit and probit models in complex settings while maintaining flexibility. Hereafter we adopt the DO multinomial probit as our predictive model, and defer to Johndrow et al. [2013] for a general discussion on details and properties of the DO multinomial class of models.

Let y_i be unordered categorical with $J = 5$ levels, and suppose $W_{i,[1:J]}$ are independent binary variables. We define

$$y_i = j \iff \{W_{ij} = 1\} \cup \{W_{ik} = 0 \quad \forall \quad k \neq j\}. \quad (8)$$

The binary variables $W_{i,[1:J]}$ have a well-known latent variable representation. In particular, $W_{ij} = 1 \iff \chi_{ij} > 0$, where $\chi_{ij} \sim N(m_{ij}, 1)$. To ensure that only one W_{ij} is equal to one, the DO model restricts the latent variables to belong to set:

$$\Omega = \cup_{j=1}^J \{\chi_i \in \mathbb{R}^J : \chi_{ij} > 0, \chi_{ik} < 0, k \neq j\}.$$

As Johndrow et al. [2013] remark, the joint probability density of χ_i 's is that of a J -variate Gaussian distribution with identity covariance that is restricted to regions of \mathbb{R}^J with one sign positive and the others negative. The categorical probabilities of class membership are easily derived as

$$\Pr(y_i = j) = \frac{(1 - \Phi(-m_{ij})) \prod_{k \neq j} \Phi(-m_{ik})}{\sum_{j=1}^J (1 - \Phi(-m_{ij})) \prod_{k \neq j} \Phi(-m_{ik})},$$

where $\Phi(\cdot)$ corresponds to the standard normal CDF. The parameters of the DO probit model can be estimated via independent binary regressions, providing substantial computational advantages over multinomial probit models (see Johndrow et al. [2013]). We model the mean of the latent Gaussian random variables as $m_{ij} = \alpha_j + \gamma_j^\top \boldsymbol{\eta}_i$, where parameter α_j can be interpreted as the baseline probability that study i is of type j whereas $\gamma_j^\top \boldsymbol{\eta}_i$ 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 proposed framework can be easily modified for joint modeling of data of many different types, e. g., the DO model for an unordered categorical outcome can be replaced by an appropriate predictive model for binary, ordered categorical, or continuous study features (for a binary example, refer to Web Appendix C). The key idea is to use the low dimensional vectors $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_n$ in all subsequent parts of the model where one seeks to link intensities $\log \mu_1, \dots, \log \mu_n$ with other variables of interest. We finally remark that we do not choose $\boldsymbol{\theta}_i$ for this task because this vector has a much bigger dimension than that of $\boldsymbol{\eta}_i$, and its inclusion in the predictive model for study type would introduce unnecessary complications in posterior update of $\boldsymbol{\theta}_i$ while also increasing the dimensionality of vectors $\{\gamma_j\}_{j=1}^J$.

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 ($\alpha_j, \gamma_j, \boldsymbol{\beta}, \boldsymbol{\Lambda}, \boldsymbol{\Sigma}$) 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.

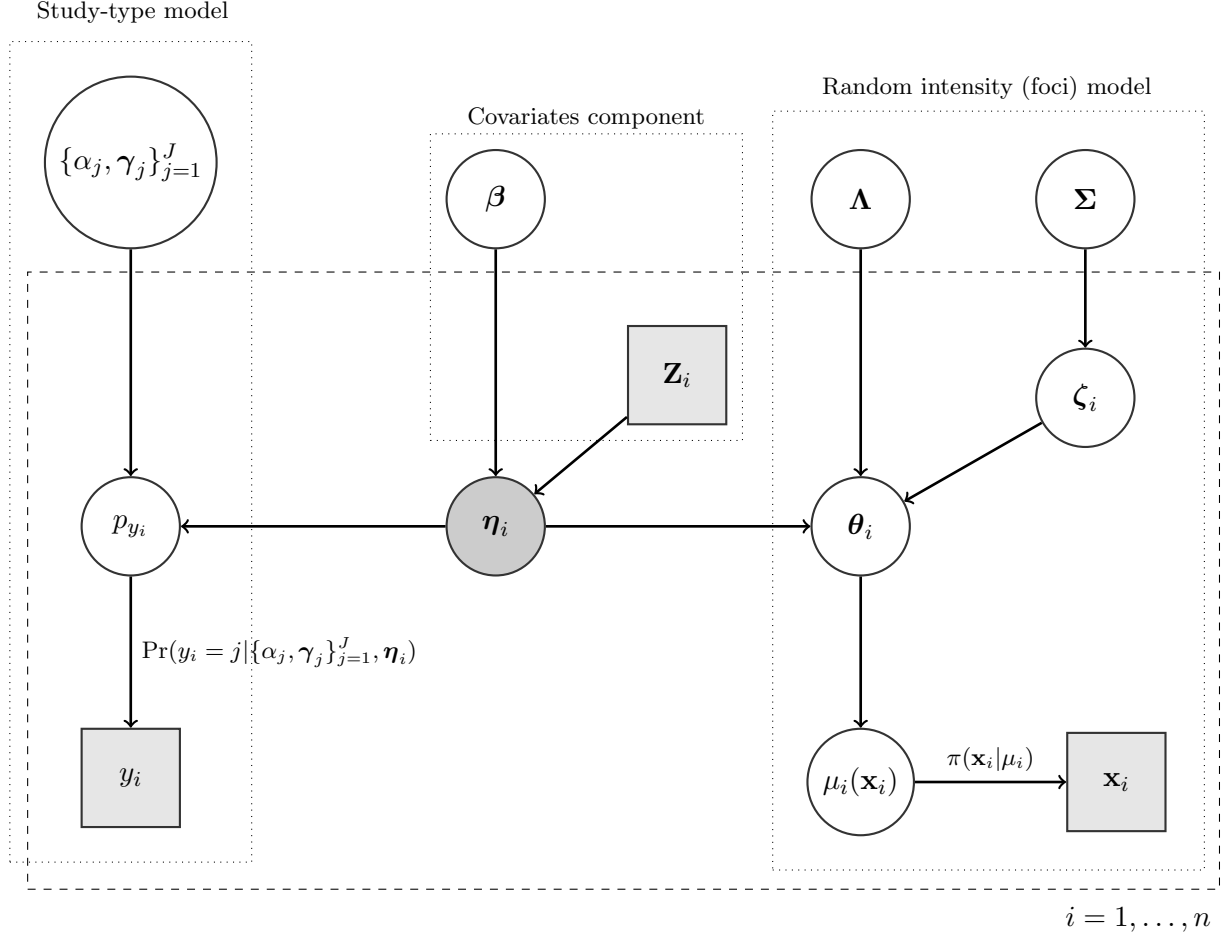


Figure 1: Graphical representation of the probabilistic mechanism generating data $\{\mathbf{x}_i, y_i\}$, $i = 1, \dots, n$, under the spatial Bayesian latent factor model. Shaded squares represent observed quantities and circles represent unknowns. The circle denoting the vector of latent factors is darkened. Note that the study type, y_i , is conditionally independent of all other nodes in the “random intensity model” given the latent variables $\boldsymbol{\eta}_i$.

2.3 fPCA-analogue construction

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

$$\log \mu_i(\boldsymbol{\nu}) = \sum_{l=1}^k \eta_{il} \phi_l(\boldsymbol{\nu}) + r_i(\boldsymbol{\nu}), \quad (9)$$

with

$$\phi_l(\boldsymbol{\nu}) = \sum_{m=1}^p \lambda_{lm} b_m(\boldsymbol{\nu}) \quad \text{and} \quad r_i(\boldsymbol{\nu}) = \sum_{m=1}^p \zeta_{im} b_m(\boldsymbol{\nu}), \quad (10)$$

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-Loève 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 k th largest eigenvalue. The empirical version (i.e., with the coefficients computed from a sample) is known as fPCA. Decomposition (9), without $r_i(\boldsymbol{\nu})$, is analogous to a truncated fPCA representation of $\log \mu_i(\boldsymbol{\nu})$, however bases $\{\phi_l\}_{l=1}^k$ are no longer mutually orthogonal within our construction. Orthogonality enhances interpretability of the elements of the decomposition, but this is not a primary concern in our application because we view the latent factorisation only as a vehicle to link the intensities with other variables. To highlight this difference with fPCA, we refer to $\{\phi_l\}_{l=1}^k$ as a dictionary.

The size k is chosen adaptively during posterior computation and the elements of the dictionary depend on the modeling of $\boldsymbol{\Lambda}$. A discussion on prior specification for the model parameters is presented in Web Appendix A of the Supplementary Materials.

3 Neuroimaging meta-analysis application

In this Section, we illustrate our approach on a meta-analysis of emotions first reported in Kober et al. [2008]. The dataset consists of 62 publications on five emotions (sad, happy, fear, anger, disgust), for a total of 187 studies and 938 foci shown in Web Figure 1. For each study, we also observe modality (fMRI/PET), inference method (fixed vs. random effects), p -value correction (corrected vs. uncorrected), the number of subjects scanned, and the type of stimulus (auditory, visual, recall, imagery, visual and auditory, olfaction), for a total of $r = 5$ covariates. Table 1 lists some summary statistics of this dataset. It is important to remark that the assumption of independence between studies (Equation (1)) might be violated here, for example if multiple experiments were run on the same subjects, and this could potentially influence our findings.

Given the sparsity of this dataset, it becomes crucial to borrow information across the population

Table 1: Data summaries.

(a) Descriptive statistics.

	Min	Median	Mean	Max.
Studies per publication	1	2	3.02	9
Foci per study	1	4	5.02	22
Subjects per study	5	12	13.56	40

(b) For each emotion type: total number of foci, total number of studies, frequency of modality (fMRI/PET), inference method (Fixed/Random effects), corrected vs. uncorrected thresholds used, and type of stimulus.

	Emotions					Total
	Sad	Happy	Fear	Anger	Disgust	
Total number of foci	220	92	264	99	263	938
Number of studies	33	27	62	22	43	187
fMRI	17	12	53	15	39	136
PET	16	15	9	7	4	51
Fixed	18	14	30	10	12	84
Random	15	13	32	12	31	103
Corrected p -values	9	5	12	3	14	43
Uncorrected p -values	24	22	50	19	29	144
Auditory	1	1	5	1	2	10
Visual	20	17	52	16	37	142
Recall	10	8	3	1	1	23
Imagery	2	—	—	4	2	8
Visual & auditory	—	1	2	—	—	3
Olfaction	—	—	—	—	1	1

of intensities to improve inferences. Specifically, the model allows borrowing strength across the different studies in estimating their intensity functions in that the low dimensional dictionary functions $\{\tilde{\phi}_m\}$, their number, and the random intercept $r_i(\boldsymbol{\nu})$ are learnt by pooling information from all studies.

We assigned a $\text{Gamma}(1, 0.3)$ prior distribution with mean $1/3$ to the diagonal elements of $\boldsymbol{\Sigma}^{-1}$. We set $p = 424$ Gaussian kernels with bandwidth $b = 0.002$. Kernels were placed on axial slices roughly 8-9 mm apart, at $z = \{-36, -28, -19, -10, -2, 7, 16\}$ mm and, within each slice, were equally spaced by forming a grid of 8×8 knots along the (x, y) direction. We used a standard brain mask with 2 mm^3 voxels and dimensions $91 \times 109 \times 32$. Kernels falling outside this mask were discarded. We performed a sensitivity analysis on the priors for σ_j^{-2} , the factor loadings, on p and b , and found no substantive differences. 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 $\text{Poisson}(30)$ 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 15,000 iterations, with the first 5,000 samples discarded as a burn-in and collecting every 25th 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 log intensity function $\mu_i(\boldsymbol{\nu})$ at a variety of voxels and for different studies. The sampler appeared to converge rapidly and mix efficiently (Web Figure 2). Further, we used the Gelman-Rubin statistics [Gelman and Rubin, 1992] to assess convergence on the number of latent factors, k . The mean of the potential scale reduction factor is 1 with an upper 0.975 quantile of 1.02. Thus, the number of iterations and burn-in appears to be sufficient.

Figure 2 shows the estimated posterior mean group intensity for the five emotion types. The group intensity at iteration t , $\hat{\mu}_g^t$, is obtained by averaging the basis function coefficients for studies that belong to the group, $\hat{\mu}_g^t(\boldsymbol{\nu}) = \exp\{\mathbf{b}(\boldsymbol{\nu})^\top \hat{\boldsymbol{\theta}}_g^t\}$, where $\hat{\boldsymbol{\theta}}_g^t = \frac{1}{\text{Card}(g)} \sum_{i \in g} \hat{\boldsymbol{\theta}}_i^t$ and $\text{Card}(g)$ is the cardinality of group g . These maps reflect the degree of consistency with which a region is activated by either emotion. All emotions show aggregation of foci in the amygdalae (the brighter

regions at axial slices $z = -27, -19, -11$ mm), although to varying degrees. The amygdalae are almond-shaped structures in the brain of known importance in emotion processing. The estimated intensity is larger in the amygdalae for disgust and fear, which are the two emotion types with more foci and studies. It is also of interest to examine the dictionary elements $\{\phi_l\}_{l=1}^k$. The interpretation of these 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 $k = 5$ with 95% credible interval $[4, 6]$. Figure 3 shows the first five elements of the dictionary $\{\phi_l\}_{l=1}^5$ (rows) at several axial slices (columns). 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 successive elements progressively shrunk toward zero. This effect is induced by a shrinkage prior on the factor loadings (see Web Appendix A). At every axial slice, the first dictionary element recovers the principal patterns of activation we observed in Figure 2, namely activation in the amygdalae. Subsequent dictionary elements are harder to interpret and of more marginal effect.

Figure 4 shows the covariate coefficients maps as of Equation (6). In particular, $\text{map } \sum_{l=2}^k \beta_{jl} \phi_l(\boldsymbol{\nu})$ quantifies the expected difference in the mean log intensity function at voxel $\boldsymbol{\nu}$ for a one-unit change in the value of covariate j , with all other quantities being equal. Notice that the first element of this sum, the intercept, is removed for illustrative purposes, and all maps are plotted on the same color scale. It appears that the covariates with strongest effect on the mean log intensity function are modality (whether the study is PET or fMRI) and p -value correction for multiple hypothesis testing. As expected, failing to correct for multiplicity results in a higher mean log intensity function particularly in the amygdalae, thus one expects here more foci than those reported by a study that controls for multiple testing. The map for modality (first row in Figure 4) is qualitatively similar to that of p -value correction (last row in Figure 4), while other maps are shrunk to zero and do not seem indicate a strong covariate effect.

For reverse inference, we split the data into a training set, for which both foci and study type are

retained for the analysis, and a testing set (80%), for which the foci only are retained, and we test the predictive accuracy of our model using the DO multinomial probit model of Section 2.2. We compare our 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 computationally efficient, but ignores the spatial dependence in the activation maps, leading to biased predictive probabilities of the class membership. Table 2 shows the out-of-sample classification rates based on our model as well as those based on the MKDA using the NBC. The simple average of correct classification rates over emotions equals 0.26 for our model and 0.25 for MKDA + NBC. Both methods tend to classify studies in the test set as fear, which is the most represented emotion in our dataset. While MKDA + NBC does a better job in correctly classifying fear studies, we do better than MKDA in correctly classifying the other emotions, in particular happiness and sadness. In general, however, the sparsity of this dataset and the limited number of studies make classification a challenging task, and both methods only slightly go above the random classification chance of 0.20.

We also tested our model on a meta-analysis dataset of emotion and executive control studies (Web Appendix C). There is substantial convergence about the systems broadly involved in each domain, and though they interact, cognitive control and emotion are associated with distinct large-scale networks. With this richer dataset, it becomes more evident that taking into account the spatial information in the data helps achieving better predictive performance over the MUA. Finally, simulations and sensitivity analyses are reported in the Supplementary Materials (Web Appendix D).

Table 2: Out-of-sample classification rates. The average correct classification rate is 0.264 for the spatial Bayesian latent factor model (SBLFM) and 0.258 for MKDA + NBC.

	Truth	Correct Classification Rates				
		Anger	Disgust	Fear	Happy	Sad
SBLFM	Anger	0.05	0.12	0.56	0.09	0.18
	Disgust	0.04	0.26	0.53	0.06	0.11
	Fear	0.06	0.15	0.56	0.10	0.13
	Happy	0.08	0.11	0.33	0.18	0.30
	Sad	0.09	0.08	0.43	0.13	0.27
MKDA + NBC	Anger	0.00	0.00	0.91	0.00	0.09
	Disgust	0.03	0.21	0.58	0.09	0.09
	Fear	0.02	0.03	0.92	0.03	0.00
	Happy	0.00	0.00	0.88	0.04	0.08
	Sad	0.12	0.03	0.73	0.00	0.12

4 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 meta-analysis. 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 an imaging predictor, the log intensity function, with an unordered categorical response, the study type, within a framework of 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 DO multinomial probit model can be replaced by an appropriate predictive model for binary, ordered categorical, or continuous study features.

There are a couple of limitations affecting our approach. First, our model is suited to studying

only activations or only deactivations, but not both simultaneously. Second, as evident in Equation (1), our model treats the studies as independent and does not account for within-experiment and within-group effects on the results. Within-experiment effects occur when studies reporting multiple foci close together in a given activation area of the brain may have a stronger influence on inference and prediction than studies reporting a single focus. Within-group effects occur when the same group of subjects is used to investigate multiple similar tasks, usually in the same scanning session, thus the resulting activation patterns can not be considered as independent observations. While highly simplifying the mathematical layout and computation for our model, the assumption that studies are truly independent might often be violated in practice. Both limitations above will be investigated in future research.

Another 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

For assistance in collecting, coding, and sharing the activation coordinates used in the meta-analyses, we would like to thank Derek Nee, Simon Eickhoff, Claudia Eickhoff (Rottschy), Brian Gold, John Jonides, Ajay Satpute, Kristen Lindquist, Eliza Bliss-Moreau, along with co-authors of the meta-analysis papers from which the coordinates were drawn. This work was partially funded by Award Number 100309/Z/12/Z from the Wellcome Trust. Dr. Johnson, Dr. Nichols, and Dr. Wager were partially supported by NIH Grant number 5-R01-NS-075066. Dr. Wager was further supported by NIH Grant number R01-DA-035484. Dr. Feldman Barrett was supported by a US National Institute on Aging grant (R01AG030311) and a contract from the US Army Research Institute for the Behavioral and Social Sciences (W5J9CQ-11-C-0046). This work represents the views of the authors and not necessarily those of the NIH, the Wellcome Trust, or the Department of the Army.

References

- Gerhard Arminger and Bengt O Muthén. A Bayesian approach to nonlinear latent variable models using the Gibbs sampler and the Metropolis-Hastings algorithm. *Psychometrika*, 63(3):271–300, 1998.
- Gregory G. Brown, Joanna E. Perthen, Thomas T. Liu, and Richard B. Buxton. A primer on functional magnetic resonance imaging. *Neuropsychology Review*, 17:107–125, 2007.
- J. Carp. The secret lives of experiments: Methods reporting in the fmri literature. *NeuroImage*, 63(1):289–300, 2012.
- D. R. Cox. Some statistical methods connected with series of events. *Journal of the Royal Statistical Society Series B*, 17(2):129–164, 1955.
- S. B. Eickhoff, A. Laird, C. Grefkes, L. E. Wang, K. Zilles, and P. T. Fox. 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, 2009.
- A. Gelman and D. B. Rubin. Inference from iterative simulation using multiple sequences. *Statistical Science*, 7:457472, 1992.
- James E Johndrow, David B Dunson, and Kristian Lum. Diagonal Orthant Multinomial Probit Models. *Proceedings of the Sixteenth International Conference on Artificial Intelligence and Statistics (AISTATS)*, 31:29–38, 2013. URL <http://jmlr.org/proceedings/papers/v31/johndrow13a.html>.
- Jian Kang, Timothy D Johnson, Thomas E Nichols, and Tor D Wager. Meta-analysis of functional neuroimaging data via Bayesian spatial point processes. *Journal of the American Statistical Association*, 106(493):124–134, March 2011. ISSN 0162-1459. doi: 10.1198/jasa.2011.ap09735. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3119536&tool=pmcentrez&rendertype=abstract>.

- 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.
- H. Kober, L. F. Barrett, J. Joseph, E. Bliss-Moreau, K. Lindquist, and T. D. Wager. Functional grouping and cortical-subcortical interactions in emotion: A meta- analysis of neuroimaging studies. *NeuroImage*, 42:998–1031, 2008.
- D. J. C. Mackay. Introduction to Gaussian processes. In C. M. Bishop, editor, *Neural Networks and Machine Learning*, pages 133–165. Springer, 1998.
- J. Møller, A. R. Syversveen, and R. P. Waagepetersen. Log gaussian cox processes. *Scandinavian Journal of Statistics*, 25:451–482, 1998.
- Silvia Montagna, Surya T Tokdar, Brian Neelon, and David B Dunson. Bayesian latent factor regression for functional and longitudinal data. *Biometrics*, 68(4):1064–73, 2012. ISSN 1541-0420. doi: 10.1111/j.1541-0420.2012.01788.x. URL <http://www.ncbi.nlm.nih.gov/pubmed/23005895>.
- 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.
- T. E. Nichols and A. P. Holmes. Nonparametric permutation tests for functional neuroimaging: a primer with examples. *Human Brain Mapping*, 15:1–25, 2001.
- J. J. Pekar. A brief introduction to functional mri. *IEEE Engineering in Medicine and Biology Magazine*, 25:24–26, 2006.
- Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.
- H. Rue and L. Held. *Gaussian Markov Random Fields*. Chapman and Hall/CRC, 2005.

- Pantelis. Samartsidis, Silvia. Montagna, Thomas. E. Nichols, and Timothy. D. Johnson. The coordinate-based meta-analysis of neuroimaging data. *ArXiv e-prints*, October 2016.
- C. R. Tench, R. Tanasescu, D. P. Auer, W. J. Cottam, and C. S. Constantinescu. Coordinate based meta-analysis of functional neuroimaging data using activation likelihood estimation; full width half max and group comparisons. *PLoS ONE*, 9:111, 2014.
- J. Thirion, P. Pinel, S. Meriaux, A. Roche, S. Dehaene, and J.-B. Poline. Analysis of a large fmri cohort: Statistical and methodological issues for group analyses. *NeuroImage*, 35:105–120, 2007.
- P. E. Turkeltaub, G. F. Eden, K. M. Jones, and T. A. Zeffiro. Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage*, 16:765–780, 2002.
- T. D. Wager, J. Kang, T. D. Johnson, T. E. Nichols, A. B. Satpute, and L. F. Barrett. A bayesian model of category-specific emotional brain responses. *PLOS Computational Biology*, 11(4):e1004066, 2015. doi: doi:10.1371/journal.pcbi.1004066.
- Tor D. Wager, M. Lindquist, and Lauren Kaplan. Meta-analysis of functional neuroimaging data: current and future directions. *Social Cognitive and Affective Neuroscience*, 2(2):150–158, 2007.
- R. L. Wolpert and K. Ickstadt. Poisson/gamma random field models for spatial statistics. *Biometrika*, 85:251–267, 1998.
- ATA Wood and G. Chan. Simulation of stationary gaussian processes in $[0, 1]^d$. *Journal of Computational and Graphical Statistics*, 3:409–432, 1994.
- K. J. Worsley, S. Marrett, P. Neelin, A. C. Vandal, K. J. Friston, and A. C. Evans. A unified statistical approach for determining significant signals in images of cerebral activation. *Human Brain Mapping*, 4:58–73, 1996.
- Tal Yarkoni, Russell a Poldrack, Thomas E Nichols, David C Van Essen, and Tor D Wager. Large-scale automated synthesis of human functional neuroimaging data. *Nature methods*, 8(8):665–70, August 2011. ISSN 1548-7105. doi: 10.1038/nmeth.1635. URL <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3146590&tool=pmcentrez&rendertype=abstract>.

- Thomas B. T. Yeo, Fenna M. Krienen, Simon B. Eickhoff, Siti N. Yaakub, Peter T. Fox, Randy L. Buckner, Christopher L. Asplund, and Michael W. L. Chee. Functional specialization and flexibility in human association cortex. *Cerebral Cortex*, 25(10):3654–3672, 2015.
- Y. U. Yue, M. A. Lindquist, and J. M. Loh. Meta-analysis of functional neuroimaging data using bayesian nonparametric binary regression. *Annals of Applied Statistics*, 6:697–718, 2012.

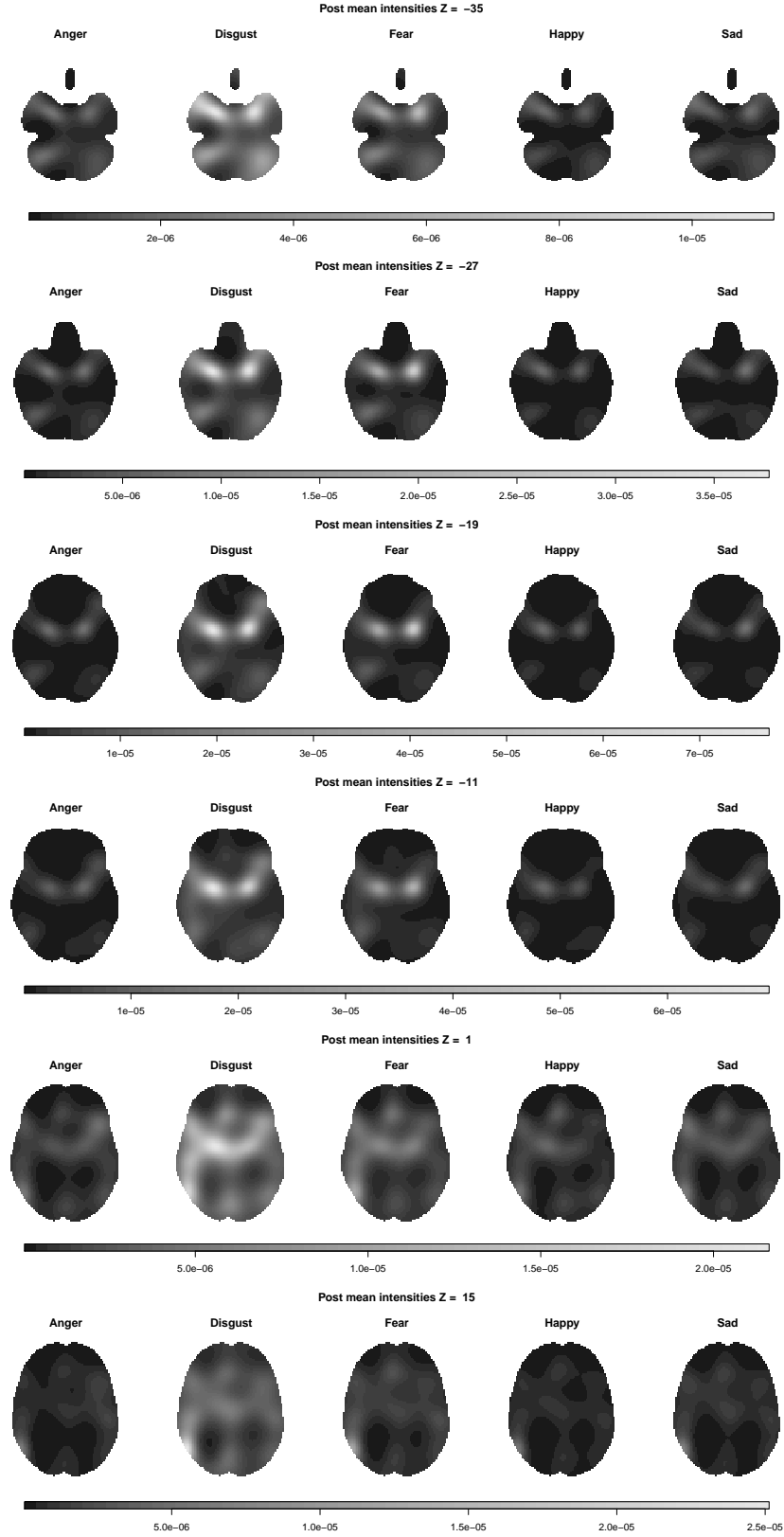


Figure 2: Posterior mean estimated intensity maps for the five emotion types. Here we only show six axial slices (rows) of the full 3D results. 24

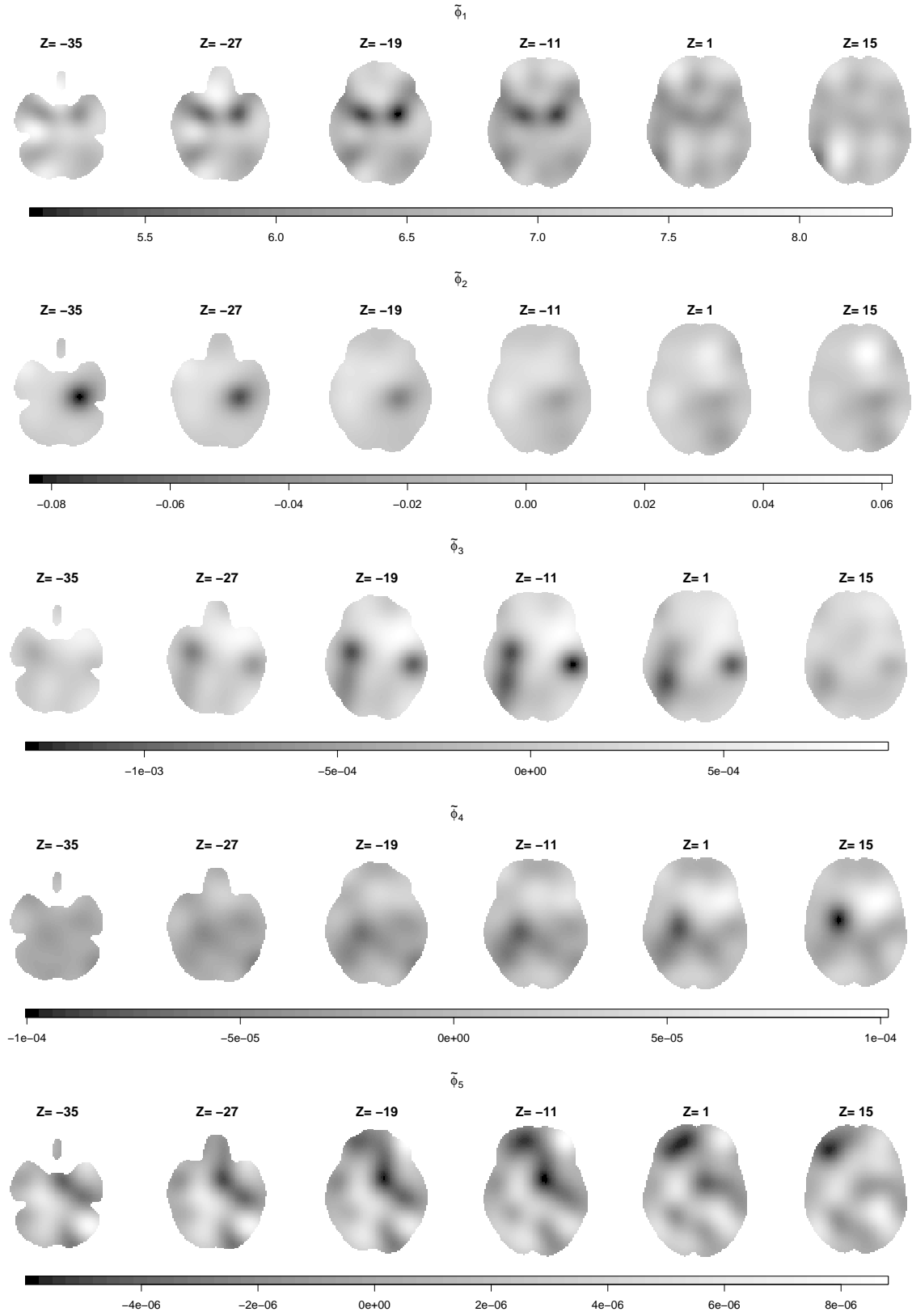


Figure 3: Learnt dictionary elements $\{\phi_l\}_{l=1}^5$ at six axial slices (columns). The estimated posterior mean number of factors is $k = 5$.

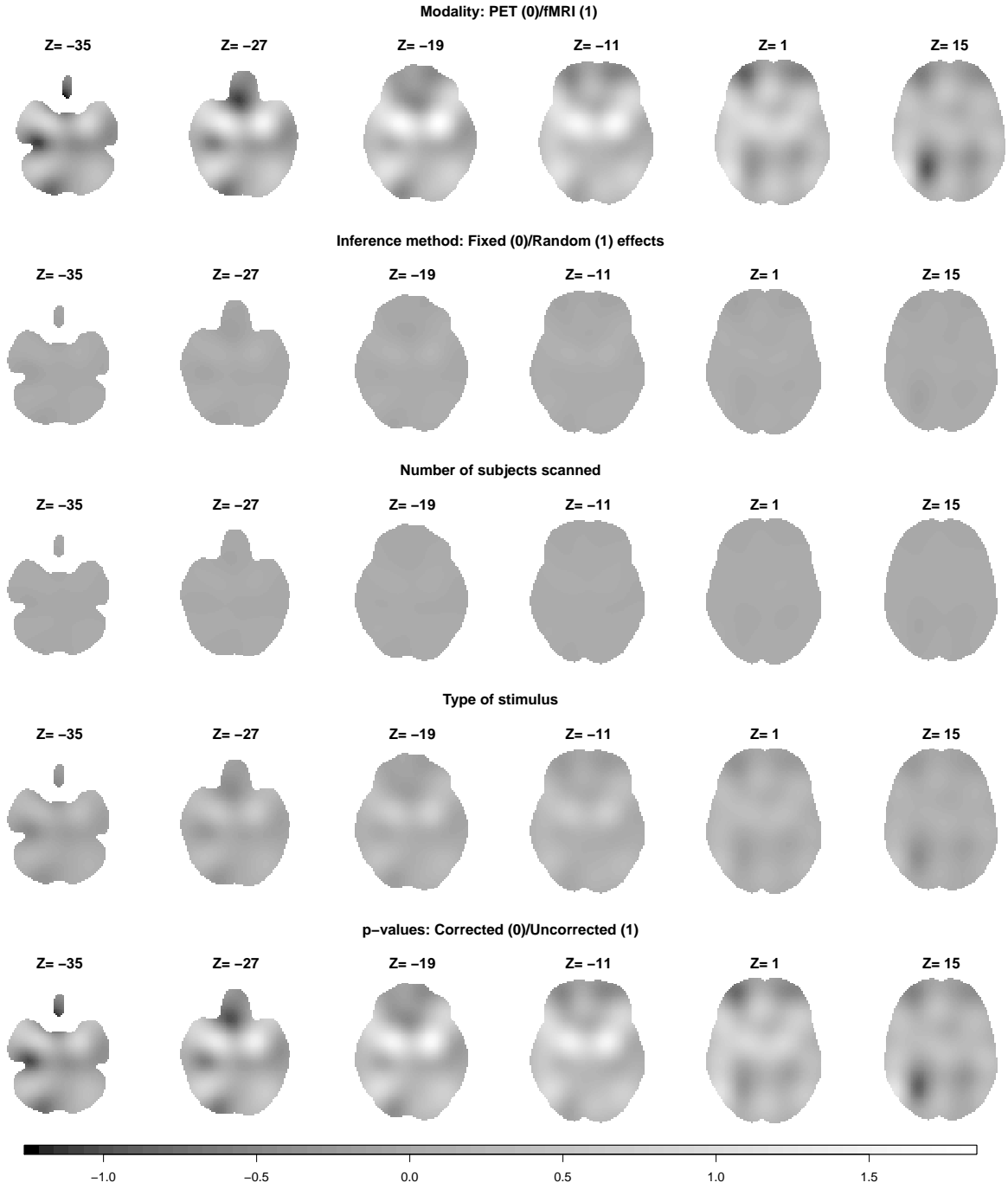


Figure 4: Estimated mean posterior covariate coefficients maps as of Equation (6) at six different axial slices. All maps are plotted on the same color scale.