

META-ANALYSIS OF FUNCTIONAL NEUROIMAGING DATA USING BAYESIAN NONPARAMETRIC BINARY REGRESSION

BY YU RYAN YUE¹, MARTIN A. LINDQUIST² AND JI MENG LOH

City University of New York, Columbia University and AT&T Labs-Research

In this work we perform a meta-analysis of neuroimaging data, consisting of locations of peak activations identified in 162 separate studies on emotion. Neuroimaging meta-analyses are typically performed using kernel-based methods. However, these methods require the width of the kernel to be set a priori and to be constant across the brain. To address these issues, we propose a fully Bayesian nonparametric binary regression method to perform neuroimaging meta-analyses. In our method, each location (or voxel) has a probability of being a peak activation, and the corresponding probability function is based on a spatially adaptive Gaussian Markov random field (GMRF). We also include parameters in the model to robustify the procedure against miscoding of the voxel response. Posterior inference is implemented using efficient MCMC algorithms extended from those introduced in Holmes and Held [*Bayesian Anal.* **1** (2006) 145–168]. Our method allows the probability function to be locally adaptive with respect to the covariates, that is, to be smooth in one region of the covariate space and wiggly or even discontinuous in another. Posterior miscoding probabilities for each of the identified voxels can also be obtained, identifying voxels that may have been falsely classified as being activated. Simulation studies and application to the emotion neuroimaging data indicate that our method is superior to standard kernel-based methods.

1. Introduction.

1.1. *Meta-analysis of neuroimaging studies.* In recent years there has been a rapid increase in the number and variety of neuroimaging studies being performed around the world. This growing body of knowledge is accompanied by a need to integrate research findings and establish consistency across labs and scanning procedures, and to identify consistently activated regions across a set of studies. Performing meta-analyses has become the primary research tool for accomplishing this goal [Wager, Lindquist and Kaplan (2007); Wager et al. (2009)]. Evaluating consistency is important because false positive rates in neuroimaging studies are likely to be higher than in many fields, as many studies do not adequately

Received January 2011; revised October 2011.

¹Supported by a PSC-CUNY Award, jointly funded by the Professional Staff Congress and the City University of New York.

²Supported in part by NSF Grant DMS-08-06088.

Key words and phrases. Binary response, data augmentation, fMRI, Gaussian Markov random fields, Markov chain Monte Carlo, meta-analysis, Spatially adaptive smoothing.

correct for multiple comparisons. Thus, some of the reported activated locations are likely to be false positives, and it is important to assess which findings have been replicated and have a higher probability of being real activations. Individual imaging studies often use very different analyses [see Lindquist (2008) for an overview], and effect sizes are only reported for a small number of activated locations, making combined effect-size maps across the brain impossible to reconstruct from published reports. Instead, meta-analysis is typically performed on the spatial coordinates of peaks of activation (peak coordinates), reported in the standard coordinate systems of the Montreal Neurologic Institute (MNI) or Talairach and Tournoux (1988), and combined across studies. These peak coordinates typically correspond to the voxel whose t -statistic takes the maximum value in a spatially coherent cluster of activation, that is, the max statistic among a set of adjacent voxels that exceed a certain threshold. This information is typically provided in most neuroimaging papers and simple transformations between the two standard spaces exist.

A typical neuroimaging meta-analysis studies the locations of peak activations from a large number of studies and seeks to identify regions of consistent activation. This is usually performed using kernel-based methods such as activation likelihood estimation [ALE; Turkeltaub et al. (2002)] or kernel density approximation [KDA; Wager, Jonides and Reading (2004)]. In both methods, maps are created for each study by convolving an indicator map, consisting of an impulse response at each study peak, with a kernel of predetermined shape and width. The resulting maps are thereafter combined across studies to create a meta-analysis map. Monte Carlo methods are used to find an appropriate threshold to test the null hypothesis that the n reported peak coordinates are uniformly distributed throughout the grey matter. A permutation distribution is computed by repeatedly generating n peaks at random locations and performing the smoothing operation to obtain a series of statistical maps under the null hypothesis that can be used to compute voxel-wise p -values. The two approaches differ in the shape of the smoothing kernel. In KDA, it is assumed to be a sphere with fixed radius, while in ALE it is a Gaussian with fixed standard deviation.

A major shortcoming of kernel-based approaches is that the width of the kernel, and thus the amount of smoothing, is fixed a priori and assumed to be constant throughout the brain. In order to address these concerns, we propose a fully Bayesian nonparametric binary regression method for performing neuroimaging meta-analysis. In our method, each location has a probability of being a peak activation, and the corresponding probability function is based on a spatially adaptive Gaussian Markov random field (GMRF). The locally adaptive features of our method allows us to better match the natural spatial resolution of the data across the brain compared to using an arbitrary chosen fixed kernel size.

In this work, a meta-analysis was performed on the results of 162 neuroimaging studies (57 PET and 105 fMRI) on emotion. The studies were all performed on healthy adults and published between 1990 and 2005. For each study, the foci

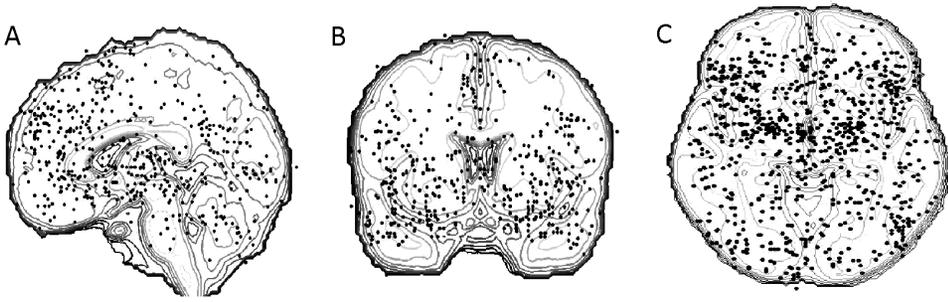


FIG. 1. Example of the raw data are shown for a representative sagittal, coronal and axial slice of the brain. Each point represents a reported activation foci in an individual study by criteria designated by that particular study. All foci are reported and plotted in the MNI brain template to allow for cross study comparisons.

of activation were included when reported as significant by the criteria designated in the individual studies. Relative decreases in activation in emotion related tasks were not analyzed. All coordinates were reported on the MNI coordinate system to allow for cross study comparisons. Together, these studies yield a data set consisting of 2478 unique peak coordinates. This data set is described in greater detail in Kober et al. (2008). Due to the relative scarcity of neuroimaging studies on a particular topic (e.g., emotion), it is standard practice in meta-analysis to combine data obtained using different imaging modalities, sample sizes and statistical analyses. This is done to ensure that the analysis has enough power to detect effects of interest. In addition, studies in Wager et al. (2008) have shown no significant difference between MRI and PET in the assessment of their functional maps and their foci of activation. Figure 1 shows the raw data for representative slices of the brain with fixed x , y and z directions, respectively. Each point in the plot represents the location of the peak of a cluster of reported activation from one of the 162 neuroimaging studies. The primary goal for analyzing this data set was to determine areas of the brain that are consistently active in studies of emotion.

1.2. *Statistical modeling for binary response.* Let Y be a random binary response variable, X a vector of covariates and $p(\mathbf{x})$ the response probability function, $p(\mathbf{x}) = \Pr(Y = 1|X = \mathbf{x})$. In the context of fMRI meta-analysis, $Y = 1$ if the voxel is reported as being a peak activation. The vector X includes the voxel location and possibly other covariates related to the patient or the study. In non-parametric binary regression, we have $p(\mathbf{x}) = H(z(\mathbf{x}))$, where H is a specified cumulative distribution function often referred to as the link function. Popular link functions are the standard logistic and standard normal cumulative density functions.

The traditional parametric approach to binary regression involves setting $z(\mathbf{x}) = \alpha + \beta^T \mathbf{x}$, with unknown parameters α and β . McCullagh and Nelder (1989) contains a comprehensive treatment of frequentist parametric methods with exponen-

tial family models, binary regression being a special case. Bayesian binary regression is well documented in, for example, [Dey, Ghosh and Mallick \(2000\)](#). In particular, [Albert and Chib \(1993\)](#) and [Holmes and Held \(2006\)](#) introduced auxiliary variable methods that provide efficient Markov chain Monte Carlo (MCMC) inference for parametric binary regression.

There is an extensive non-Bayesian literature on nonparametric regression using exponential family models, with binary regression treated as a special case. [O'Sullivan, Yandell and Raynor \(1986\)](#) estimated a single function using a penalized likelihood approach, and their work was extended to additive models by [Hastie and Tibshirani \(1990\)](#). [Gu \(1990\)](#) and [Wahba et al. \(1995\)](#) used tensor product smoothing splines to allow for interactions between variables and estimated smoothing parameters via a generalized cross-validation technique. [Loader \(1999\)](#) proposed a local likelihood approach for both univariate and bivariate nonparametric estimation and provided data-driven bandwidth estimators.

Bayesian methods for nonparametric binary regression were developed in [Wood and Kohn \(1998\)](#), [Holmes and Mallick \(2003\)](#), [Choudhuri, Ghosal and Roy \(2007\)](#), and [Trippa and Muliere \(2009\)](#). These methods are not locally adaptive, however. [Krivobokova, Crainiceanu and Kauermann \(2008\)](#) proposed an adaptive penalized spline estimator for binary regression based on quasi-likelihoods. [Wood et al. \(2008\)](#) presented a locally adaptive Bayesian estimator for binary regression by using a mixture of probit regressions where the argument of each probit regression is a thin-plate spline prior with its own smoothing parameters and the mixture weights depend on the covariates.

In fMRI meta-analysis, kernel-based smoothing techniques are typically used to identify regions of consistent activation and Monte-Carlo procedures are used to establish statistical significance. These techniques count the number of activation peaks within a radius of each local brain area and compare the observed number to a null distribution to establish significance. The kernel radius is chosen by the analyst, and kernels that match the natural spatial resolution of the data are the most statistically powerful [[Wager, Lindquist and Kaplan \(2007\)](#)]. In our method, the function $z(\cdot)$ is assumed to be a spatially adaptive Gaussian Markov random field (GMRF) with locally varying variance. The local adaptiveness of the procedure allows the probability function to be smooth in some regions and wiggly in others, depending on the data information. The need of adaptive smoothing for fMRI data has been demonstrated in [Brezger, Fahrmeir and Hennerfeind \(2007\)](#) and [Yue, Loh and Lindquist \(2010\)](#). The proposed Bayesian nonparametric binary regression method is an extension to the binary response case of methods developed in [Yue and Speckman \(2010\)](#) and [Yue and Loh \(2011\)](#). To make this procedure better suited for application to fMRI meta-analysis, we incorporate additional model parameters associated with the probabilities of voxels being miscoded. This makes the modeling more robust to possible errors in the data. The posterior inference is carried by efficient MCMC algorithms extended from those in [Holmes and Held \(2006\)](#). From the model fit we obtain a map of the probability of observing a peak

the weight given to the corresponding “•” locations. Therefore, the conditional mean of z_{jk} is a particular linear combination of the values of its neighbors, and its conditional variance is $\text{Var}(z_{jk}|\mathbf{z}_{-jk}) = 20\delta^2\gamma_{jk}^2$.

The use of γ_{jk}^2 is important for estimating activation probabilities in a fMRI meta-analysis. To identify consistently activated regions across a set of studies, we need less smoothing (large γ_{jk}^2) where there are many reported peak locations and relatively more smoothing (small γ_{jk}^2) where very few or no peaks are reported. Standard smoothing techniques (e.g., kernel smoother with fixed width) suffer from a trade-off between increased detectability and loss of information about the spatial extent and shape of the activation areas. Adaptive smoothing provided by γ_{jk}^2 can reduce such loss of information. The need of adaptive smoothing for processing fMRI imaging data was also demonstrated in Brezger, Fahrmeir and Hennerfeind (2007) and Yue, Loh and Lindquist (2010). Note that setting $\gamma_{jk}^2 \equiv 1$ makes (1) a nonadaptive GMRF on lattice, yielding a Bayesian solution for thin-plate splines [see Rue and Held (2005), section 3.4.2].

Additional priors need to be specified for γ_{jk}^2 in (1). We use independent inverse gamma priors for γ_{jk}^2 , that is, $\gamma_{jk}^{-2} \stackrel{\text{i.i.d.}}{\sim} \text{Gamma}(\nu/2, 1/2)$, $\nu > 0$. The marginal prior distribution of the increment in (1) turns out to be a Student- t distribution with ν degrees of freedom. We choose a Cauchy distribution ($\nu = 1$), which has been suggested as a default prior for robust nonparametric regression [Carter and Kohn (1996)] and sparse Bayesian learning [Tipping (2001)]. Yue and Loh (2011) and Brezger, Fahrmeir and Hennerfeind (2007) also suggested similar priors for γ_{jk}^2 in their work on adaptive spatial smoothing. Yue and Speckman (2010) and Yue, Loh and Lindquist (2010), however, assumed another spatial GMRF model for $\log(\gamma_{jk}^2)$ in a second hierarchy. Although it has been applied successfully for modeling spatial data, this two-stage GMRF prior forces the γ_{jk}^2 to be smooth and it is not suitable for estimating spatial processes with jumps or sharp peaks. Furthermore, the computation is rather complicated, precluding extensions to more flexible regression models, for example, the binary hierarchical regression model considered here.

The prior for δ^2 is often chosen to be a conjugate diffuse but proper inverse gamma prior. We, however, propose to use a half- t distribution as the prior for its square root, that is,

$$(3) \quad [\delta|\rho, S] \propto \left(1 + \frac{1}{\rho} \left(\frac{\delta}{S}\right)^2\right)^{-(\rho+1)/2}, \quad \delta > 0,$$

where ρ is the parameter of degrees of freedom and S is the scale parameter. The half- t distribution can be treated as the absolute value of a Student- t distribution centered at zero [see Psarakis and Panaretos (1990)]. Although it is not commonly used in statistics, the half- t distribution was used in objective Bayesian inference by Wiper, Girón and Pewsey (2008) and suggested for use as a default prior for

a variance component in hierarchical models [e.g., Gelman (2006); Gelman et al. (2008)]. This family includes, as special cases, the improper uniform density (if $\rho = -1$) and the proper half-Cauchy (if $\rho = 1$). Following Carvalho, Polson and Scott (2010), we use a standard half-Cauchy prior ($\rho = S = 1$) due to its heavy tail and substantial mass around zero. Although it is not conjugate, the half- t prior on δ can be written as $\delta \stackrel{D}{=} |\xi|/\theta$, where $\xi \sim N(0, 1)$ and $\theta^2 \sim \text{IG}(\rho/2, \rho S^2/2)$ [e.g., Psarakis and Panaretos (1990)]. This property enables us to develop efficient MCMC sampling schemes as shown in the Appendix.

2.2. Posterior inference. Although any cumulative distribution function (cdf) H that preserves the smoothness of \mathbf{z} may be used as a link function, here, we only consider the case in which the H can be represented as the scale mixture of mean zero normal cdf's. Two special examples are the well-known probit and logit link functions. With a specific link function, the posterior distribution of \mathbf{z} is not analytically tractable, and thus an MCMC algorithm will be used to compute the posterior distribution. The algorithm is based on the auxiliary variable method in Holmes and Held (2006) and GMRF simulation techniques in Rue and Held (2005). Briefly, the data are augmented by introducing an auxiliary variable w_i that follows a normal distribution with mean z_i and variance λ_i . The new data w_i are associated with original binary data y_i in the following way: $y_i = 1$ if $w_i > 0$ and $y_i = 0$ if $w_i \leq 0$. Then, the adaptive GMRF prior is taken on z_i and a certain prior distribution chosen for λ_i depending on the link function. The full conditional distributions for the Gibbs sampler are all easily derived and can be efficiently sampled. In the Appendix we provide the detailed MCMC algorithms for the link functions that are the probit, logit and general scale mixture of normals.

2.3. Robustification. In this section we describe how to robustify our procedure against miscoding of the response variable. Adopting the idea in Choudhuri, Ghosal and Roy (2007), we use indicator variables $\boldsymbol{\psi} = (\psi_1, \dots, \psi_n)'$ such that $\psi_i = 1$ indicates that y_i is miscoded and $\psi_i = 0$ indicates that y_i is correctly coded. In the context of fMRI meta-analysis, $\psi_i = 1$ means that y_i is either a false positive or a false negative. Since these variables cannot be observed, we treat them as unknown parameters that need to be estimated via taking priors on them. The joint posterior distribution of $(\boldsymbol{\psi}, \mathbf{z})$ is then used to obtain a robust estimate of \mathbf{z} , and also to identify the miscoded observations.

We assume that each observation has equal probability of being miscoded, independent of other observations and \mathbf{z} . Denote by r an a priori guess for the probability of an observation being miscoded. Given $(\boldsymbol{\psi}, \mathbf{z})$, the y_i 's are independent Bernoulli random variables with probability of success $(1 - \psi_i)H(z_i) + \psi_i(1 -$

$H(z_i)$). As a result, the conditional distributions of ψ_i are independent with

$$(4) \quad P(\psi_i = 1 | \mathbf{y}, \mathbf{z}) = \begin{cases} \frac{r[1 - H(z_i)]}{r[1 - H(z_i)] + (1 - r)H(z_i)}, & \text{if } y_i = 1, \\ \frac{rH(z_i)}{rH(z_i) + (1 - r)[1 - H(z_i)]}, & \text{if } y_i = 0. \end{cases}$$

Consider the probit link without any hyperprior. As shown in Section A.1, we adjust latent variables w_i for miscoding, that is, $y_i = 1$ if $\{w_i > 0, \psi_i = 0\}$ or $\{w_i < 0, \psi_i = 1\}$. Then,

$$(5) \quad (w_i | \boldsymbol{\psi}, \boldsymbol{\xi}, \boldsymbol{\eta}, \mathbf{y}) \sim \begin{cases} N(\boldsymbol{\xi} \boldsymbol{\eta}_i, 1) I(w_i > 0), & \text{if } y_i + \psi_i = 1, \\ N(\boldsymbol{\xi} \boldsymbol{\eta}_i, 1) I(w_i \leq 0), & \text{if } y_i + \psi_i \neq 1. \end{cases}$$

Hence, samples from the joint distribution $(\psi_i, w_i | \mathbf{z}, \mathbf{y})$ can be drawn by first sampling ψ_i using (4) and then sampling w_i using (5). Since the full conditional of \mathbf{z} does not depend on $\boldsymbol{\psi}$ or \mathbf{y} , the samples from the conditional distributions of the rest of the parameters can be drawn as described earlier. Note that the algorithm of this robust approach may be extended similarly to the logit link or an arbitrary symmetric link by introducing the relevant latent variables.

3. Simulation studies. We performed two different types of simulation studies to investigate the performance of our method. The first simulation is in the setting of nonparametric binary regression, where the proposed method is compared to an adaptive penalized spline model. The second simulation is in the setting of fMRI meta-analysis, where our method is compared to the kernel-based ALE method, which is commonly used in neuroimaging meta-analysis.

3.1. *Simulation I.* The true probability function is assumed to be

$$p(\mathbf{x}) = \Phi \left\{ 6 \exp \left[-\frac{5}{2} ((x_1 - 2)^2 + (x_2 - 2)^2) \right] + 3 \exp \left[-\frac{1}{10} (x_1^2 + x_2^2) \right] - 3 \right\}.$$

It is a smooth bimodal spatial surface on a 30×30 regular lattice as shown in Figure 2(a). One hundred data sets were simulated and we use the mean squared probability error (MSPE),

$$\text{MSPE} = \frac{1}{n} \sum_{i=1}^n \{p(\mathbf{x}_i) - \hat{p}(\mathbf{x}_i)\}^2,$$

to measure performance, where $\hat{p}(\cdot)$ is the estimated probability function.

The estimates obtained using our Bayesian nonparametric binary regression model are compared to those obtained using the fast adaptive penalized splines (FAPS) model in Krivobokova, Crainiceanu and Kauermann (2008). The FAPS approach models the regression function as a penalized spline with a smoothly varying smoothing parameter function which is also modeled as a penalized spline.

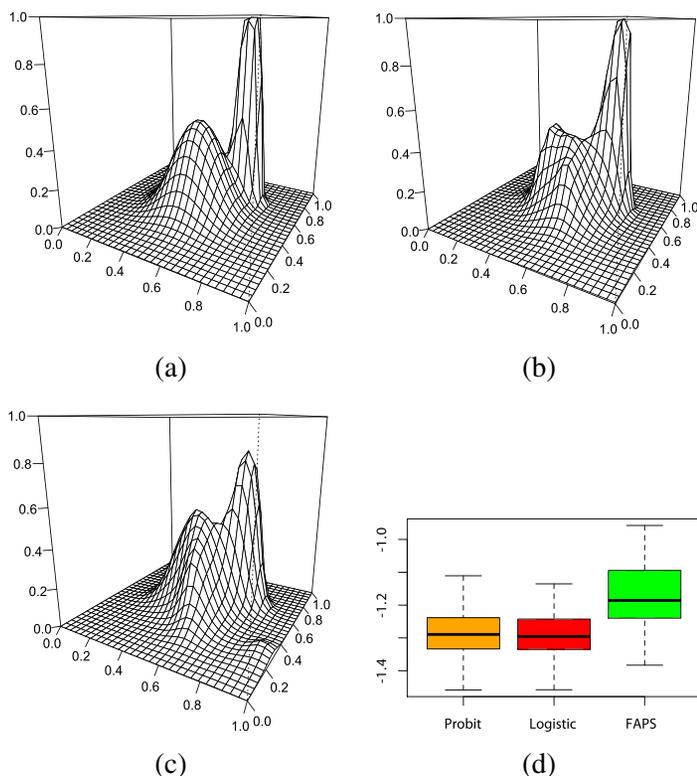


FIG. 2. *Simulation I: (a) True probability function; (b) Estimated probability function using proposed method with probit link; (c) Estimated probability function using FAPS method; (d) Mean squared probability errors using proposed method and FAPS method.*

Their method handles local smoothing of binary data as a special case. The authors showed that the FAPS estimator outperformed the penalized spline estimators in Crainiceanu et al. (2007) and Ruppert and Carroll (2000). The model can be fit using the `AdaptFit` R package.

Panels (b) and (c) in Figure 2 show typical fits for the bimodal function using our method and FAPS method, respectively. It appears that the FAPS model has difficulty capturing the sharp peak and undersmooths the flat portion as well. Figure 2(d) shows the distributions of the MSPE produced by those two methods, where the FAPS estimator is apparently outperformed. Also, in our method the two link functions yield similar performances in terms of MSPE. This is because nonparametric modeling of \mathbf{z} makes the model robust against the choice of the link function. We believe that the underperformance of FAPS stems from using slowly varying functions to model local smoothing parameters. Although they provide computational efficiency, such low-rank basis functions are unable to capture sharp changes in the function. Yue and Speckman (2010) presented similar results for

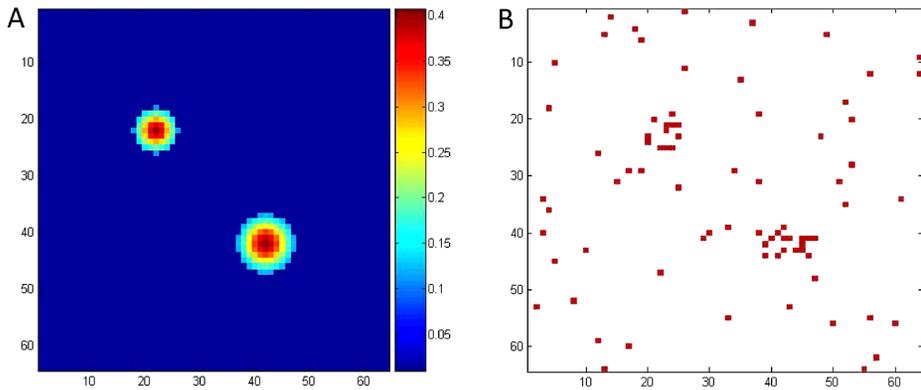


FIG. 3. (A) The probability map used to generate random activation peaks; (B) One set of simulated activation peaks.

normal response variables. Note that the robustification procedure is not required in this simulation study.

3.2. Simulation II. In the second simulation study we began by constructing a 64×64 probability map, denoted $p(x, y)$, where the value at each voxel location (x, y) represents the probability that it be recorded as a “peak coordinate” in a neuroimaging study. The probability map consisted of two circular regions of heightened probability (see Figure 3A), where the maximum probability is roughly 0.4. Voxels lying outside these two regions were set to have a constant background probability of 0.01, thus allowing for the possibility of “false positives” outside the two centers of activation. Next, the probability map was used to generate random activation peaks. The voxel at coordinate (x, y) was considered a reported peak according to a binomial distribution with probability of activation $p(x, y)$. This process was repeated 100 times and each time gives rise to simulated meta-analysis data. Figure 3B shows the data for one repetition. The data shows clear clustering around the two regions of activation, while still allowing for spurious activations in the rest of the image. This corresponds with the behavior of standard meta-analysis data (see, e.g., Figure 1).

Each of the 100 repetitions were analyzed using the kernel-based ALE method as well as our Bayesian nonparametric binary regression model. In the former, a kernel with bandwidth 10 mm full width at half maximum (FWHM) was used, as this is the standard in the field. A Monte Carlo procedure was used to determine the appropriate threshold to test the null hypothesis that the reported peak coordinates are uniformly distributed throughout the grey matter. A permutation distribution is computed by repeatedly generating peaks at random locations and performing the smoothing operation to obtain a series of statistical maps under the null hypothesis that can be used to determine which voxels had p -values below α , where α was set to 0.05 and 0.01. Regarding our Bayesian method, the robustification procedure

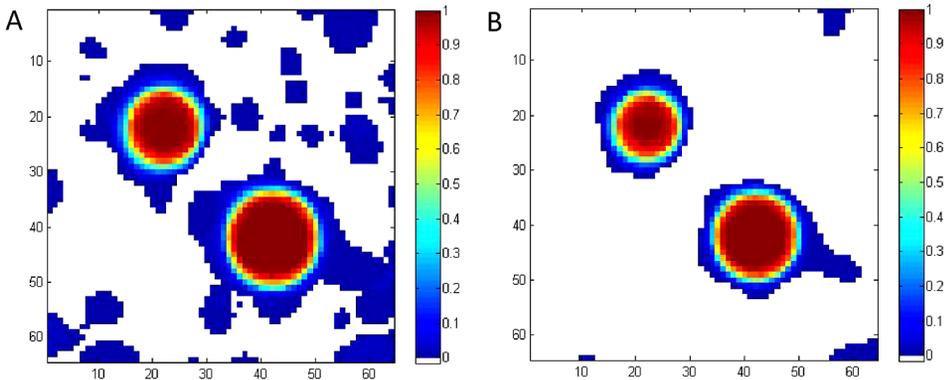


FIG. 4. Proportion of times each voxel was deemed significant at the 5% level (A) and the 1% level (B) using the ALE method.

described in Section 2.3 is implemented since we use the background probability of 0.01 to produce the false positives. To see how sensitive the results are to the use of robustification, we fit the model with prior miscoding probability $r = 0$ (no robustification), $r = 0.01$ and $r = 0.05$. Figures 4A and B show the proportion of times each voxel was deemed significant at the 5% level and the 1% level, respectively, in the 100 repetitions, when the ALE method was used. It is clear that the kernel smoother does a very good job of finding true positives, but tends to have a large number of false positives in the area immediately surrounding the activated regions. Figure 5 shows the average probability of activation in each voxel obtained using our method. The maps in the left column are not thresholded, while those in the right column are thresholded at 0.01. Apparently, our estimates are closer to the simulated probability map and produce much fewer false positives than the kernel estimates. Furthermore, our method yields fewer false positives as the value of r , the prior miscoding probability, increases, that is, the fit becomes more robust. The spatial extent of the activation region, however, is not significantly shrunk, making a strong case for the use of adaptive smoothing.

3.3. Computational performance and MCMC diagnostics. Thanks to the sparse structure of the adaptive GMRF prior used, the proposed models provide fast MCMC computation for nonparametric binary regression. To complete 5000 iterations on a 3.06 GHz Intel iMac desktop with 4GB memory, it took the probit model 9.23, 46.06 and 11.17 seconds at sample size $n = 30 \times 30$, 60×60 and 90×90 , respectively, for estimating the bimodal function in Simulation I. The logistic model is a little slower, taking 11.89, 55.83 and 138.69 seconds to finish the same amount of computations. The computing times of both models increase with sample sizes at order n , roughly. The programs were written in the FORTRAN language, making use of the LAPACK and BLAS packages.

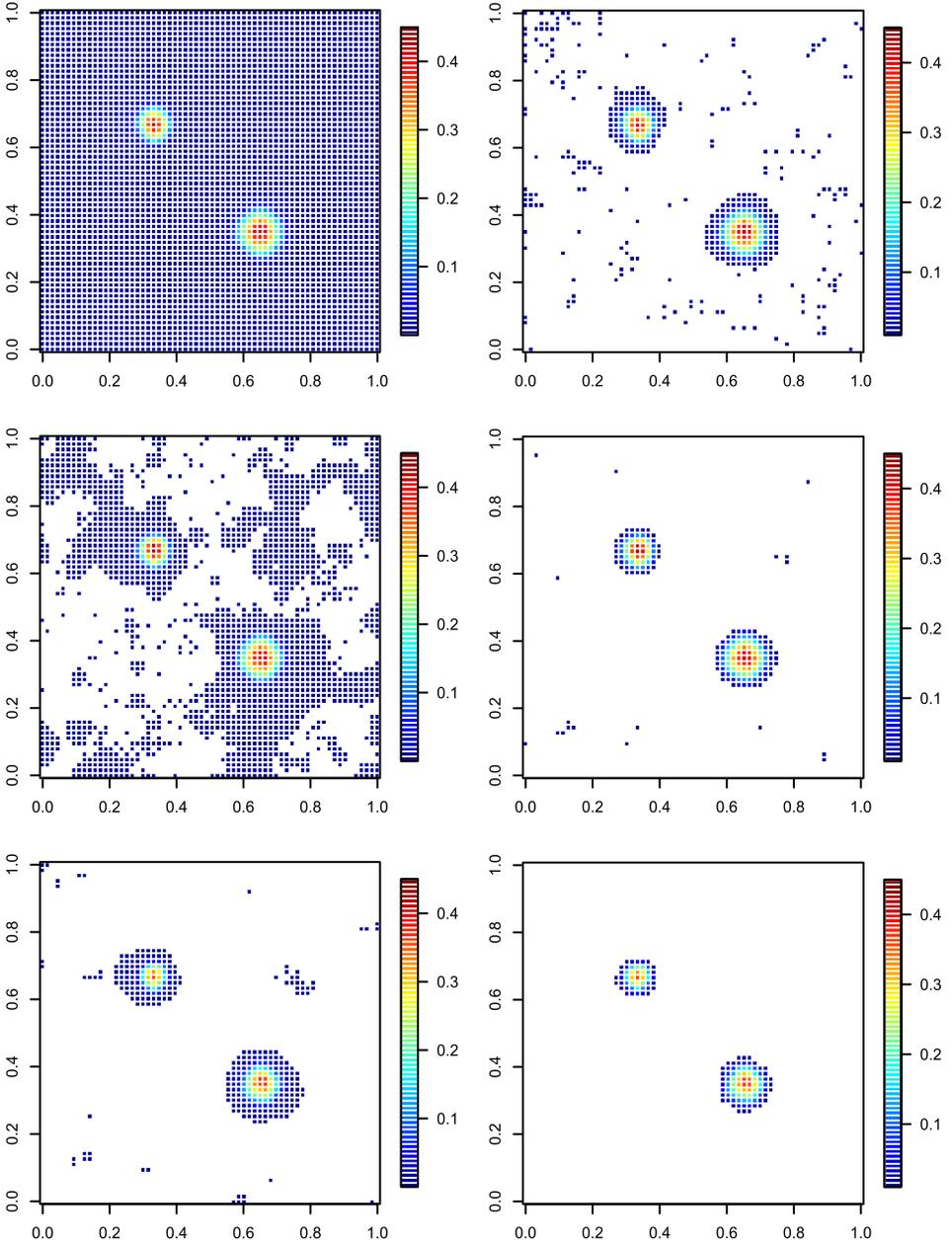


FIG. 5. Average probability of activation in each voxel obtained using the adaptive GMRF method combined with the robustification procedure under different prior miscoding probabilities: $r = 0$ (top row), $r = 0.01$ (middle row) and $r = 0.05$ (bottom row); The maps in the left column are not thresholded, while those in the right column are thresholded at 0.01.

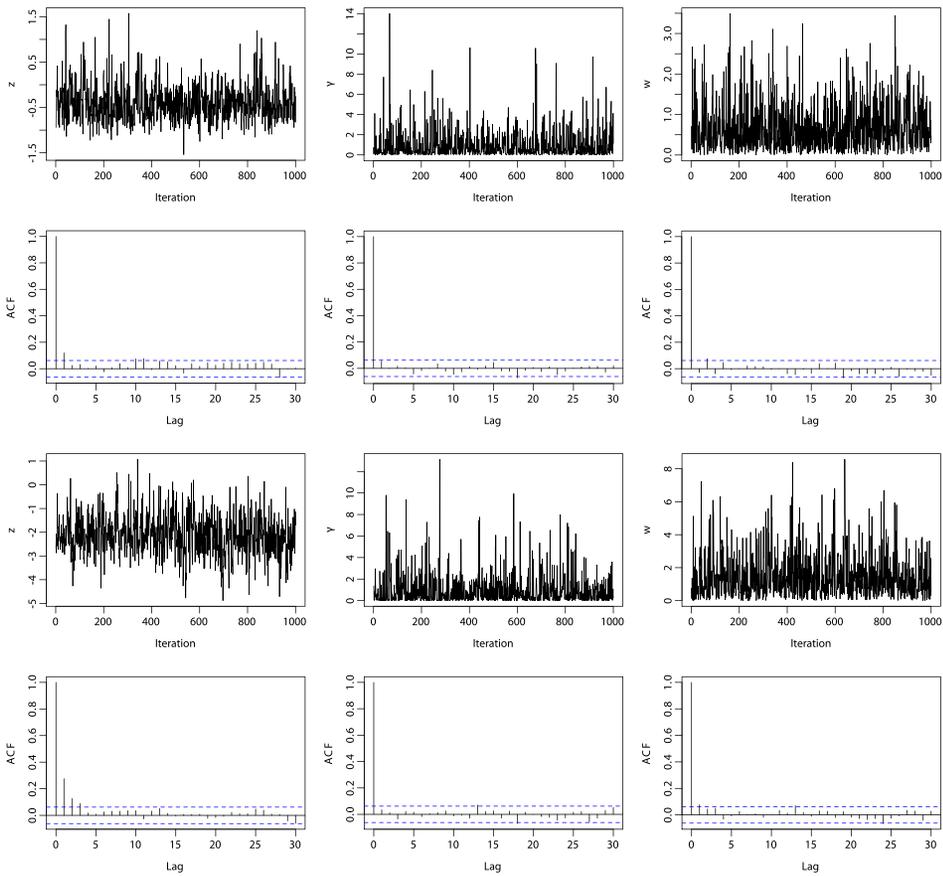


FIG. 6. Assessment of MCMC convergence for Simulation I. The top (bottom) two rows contain the typical trace plots and autocorrelation functions of the samples of variables z , γ and w from fitting a probit (logistic) model.

It is well known that the GMRF \mathbf{z} are strongly dependent on each other as well as on the auxiliary variable \mathbf{w} [see, e.g., Rue and Held (2005); Holmes and Held (2006)]. Those posterior correlations are likely to cause slow mixing in the Markov chain. To combat this issue, we sampled \mathbf{z} as a block and employed the joint updating tricks as used in Holmes and Held (2006) (see the Appendix for details). Since the computation is fast, we also suggest running a relatively large number of MCMC iterations and applying a thinning factor of ℓ by collecting samples after every ℓ iterations. In Simulation I, for instance, we found that it is sufficient to run 15,000 MCMC iterations (5000 burn-in and 10,000 sampling) with a thinning factor of 10 to obtain reliable estimates. Figure 6 shows typical trace plots and autocorrelation functions of the samples of different variables for Simulation I. As we can see, the mixing of the chain is satisfactory for both probit and logistic models.

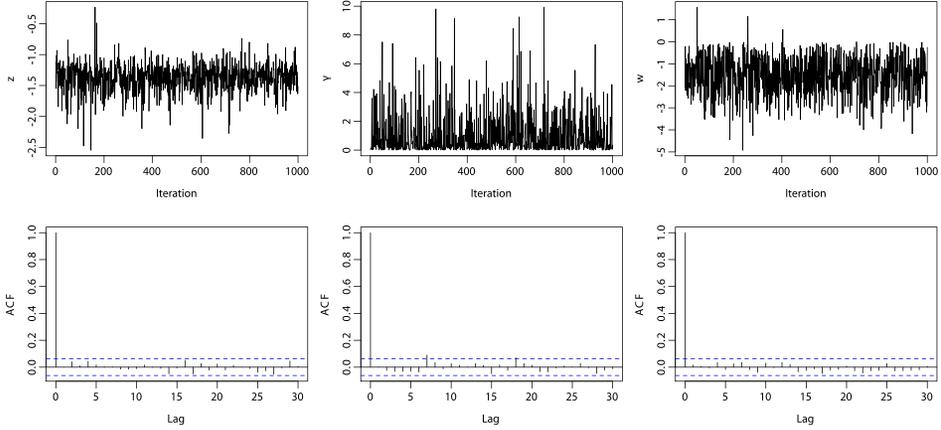


FIG. 7. Assessment of MCMC convergence for the data analysis. The top row contains the typical trace plots of the samples of variables z , γ and w ; the bottom row contains the corresponding autocorrelation functions.

4. Data analysis. We describe here the results of our meta-analysis of the fMRI data. As mentioned before, the data consists of the coordinates of 2478 peaks representing the locations of voxel activations, collected from 162 neuroimaging studies. The raw data consists of a three-dimensional image with dimensions $91 \times 109 \times 91$ whose elements took the value 1 if an activation had been reported at that voxel and 0 otherwise. Figure 1 shows the raw data for a representative slice of the brain with fixed x , y and z directions, respectively.

The binary nature of the meta-analysis data makes it an ideal candidate for our Bayesian nonparametric binary regression method. As our method is currently only implemented in two dimensions, we fit our method slice-wise across the brain for each orientation (i.e., for the fixed x , y and z direction). Prior to performing our method on a slice, we applied smoothing in the fixed direction by including all activations located within 10 mm of the slice of interest.

In our simulation studies (Section 3), we found that the binary regression model is not sensitive to the choice of link function. We therefore fit a probit model to the data for computational efficiency. To make our estimation robust against false positives, we incorporated the robustification procedure (Section 4) in the model with prior miscoding probability $r = 0.01$ for every voxel. Due to the high dimension of the data, the MCMC was run for 60,000 iterations with 10,000 burn-in and a thinning factor of 50 iterations, resulting in posterior samples of size 1000. The Markov chains mix well as shown in Figure 7.

Once the Bayesian binary regression model was fit, posterior probability maps were obtained indicating the probability of being a location of peak activation across the brain. Regions with probability values higher than 0.3 were color-coded and superimposed onto an anatomical reference image. The relatively low threshold is indicative of the dispersion of foci locations in the data. Figure 8 shows

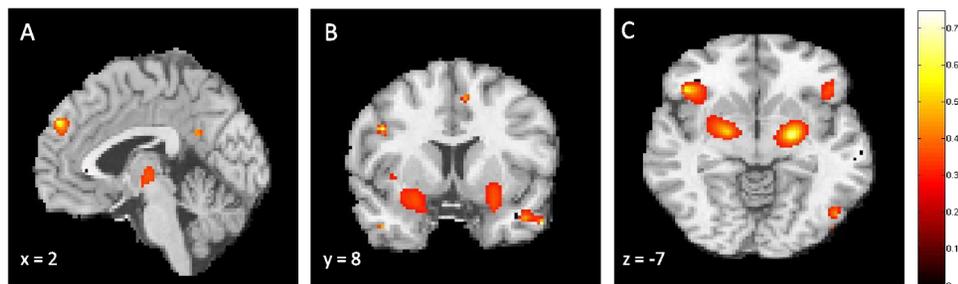


FIG. 8. Thresholded posterior probability maps are shown for the sagittal, coronal and axial slice of the brain depicted in Figure 1. Regions with posterior probability of observing a peak activation higher than 0.30 are color-coded.

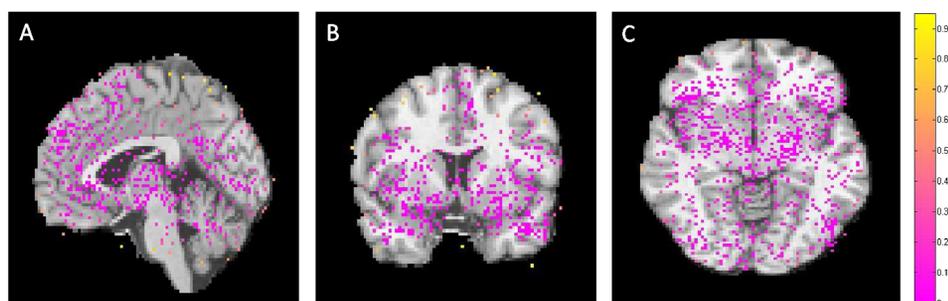


FIG. 9. Miscoding probabilities are shown for the sagittal, coronal and axial slice of the brain depicted in Figure 1. Points with posterior miscoding probability higher than 0.10 are color-coded.

results for the three slices described above. Key regions of activation observed in the figure include the thalamus (8A), amygdala (8B) and the ventral striatum (8C). These regions are known to be associated with emotion, and were also indicated as active when using kernel-based methods [see Kober et al. (2008)]. It should be noted we obtain the same regions of activation as Kober et al. (2008), but with significantly smaller spatial extent. This is consistent with our simulation study, which shows how the kernel-based methods tend to overestimate the extent of activation. Finally, Figure 9 shows the posterior miscoding probabilities (thresholded at 0.10) for the same three slices. High miscoding probabilities indicate points that were deemed to be spurious activations and therefore given lower weights when calculating the posterior activation probabilities. Based on their locations, it appears that our method is providing an effective means of downweighting false activations.

To see if the adaptive smoothing is preferred to the ordinary smoothing in this neuroimaging example, we conducted a test on $H_0: \gamma_{jk} = 1$ using the deviance information criterion (DIC) introduced by Spiegelhalter et al. (2002). More specifically, we first fitted to our imaging data the proposed adaptive GMRF model and a (nonadaptive) Bayesian thin-plate spline model (by fixing all γ_{jk} to be 1),

TABLE 1
*DIC scores of both adaptive and nonadaptive models for the
 fixed x , y and z orientation.*

Orientation	x	y	z
Adaptive	9918.372	8216.255	9917.209
Nonadaptive	10,090.460	9512.016	9947.377

and saved the MCMC posterior samples of both models. Then, we define the deviance as $D(\phi) = -2 \log(p(\mathbf{y}|\phi))$, where $p(\mathbf{y}|\phi)$ is the likelihood function and ϕ are unknown parameters of the model. The DIC score is finally estimated using $DIC = 2\bar{D} - D(\bar{\phi})$, where \bar{D} is calculated as the average of $D(\phi)$ over the samples of ϕ , and $D(\bar{\phi})$ as the value of D evaluated at the average of the samples of ϕ . The model with smaller DIC should be in favor. Table 1 shows the DIC scores of the two models for the fixed x , y and z orientations, where the adaptive model is preferred in every scenario.

5. Discussion. We developed a fully Bayesian method for nonparametric binary regression and, together with a robustification procedure, applied it to meta-analysis in fMRI studies. Our analysis identified activated regions of the brain that are known to be associated with emotion. While similar regions were also identified in other meta-analyses such as [Kober et al. \(2008\)](#) that use kernel-based methods, our method has several advantages over such approaches as follows. The adaptive GMRF used in our model better matches the natural spatial resolution of the data across the brain compared to using an arbitrary chosen fixed kernel size. This allows us to avoid the problem of overestimating regions of activation apparent in kernel-based methods. The Bayesian nature of our method allows for the construction of posterior probability maps indicating the probability of observing a peak activation in response to the paradigm across the brain. This is in contrast to kernel methods which simply state that more peaks lie near the voxel than expected by chance. It should be noted that recently a Bayesian spatial hierarchical model using a marked independent cluster process [[Kang et al. \(2011\)](#)] was introduced for dealing with neuroimaging meta-analysis. In future work we will look at comparing this method with the nonparametric binary regression approach suggested in this paper. Finally, our procedure provides estimates of miscoding probabilities which can help to identify regions that may have been incorrectly tagged as being activated. This is another feature not provided by kernel-based methods.

It is important to note that in this work the model setup assumes that the input data is two dimensional. Such 2D smoothing serves a useful purpose, as fMRI data are often analyzed either slice-wise or using cortical surface-based techniques [[Dale, Fischl and Sereno \(1999\)](#); [Fischl, Sereno and Dale \(1999\)](#)]. In reality, however, fMRI data are three dimensional in space. Therefore, it may ultimately be more appropriate to smooth the three spatial dimensions directly. We

are actually working on such an extension of our current approach. The main computational constraint stems from inverting a large precision matrix, which is of $91 \times 109 \times 91 = 902,629$ dimensions in our neuroimaging example. We thus need a practical 3D GMRF, but the construction is nontrivial. One possible solution is to obtain a highly sparse precision matrix by discretizing a 3D Laplacian operator with proper boundary conditions as we did in the 2D case. To achieve more computational efficiency, we may use a novel Bayesian inference tool similar to that introduced in [Rue, Martino and Chopin \(2009\)](#) rather than MCMC.

As shown in the simulation studies, the results obtained by our method are somewhat sensitive to the prior miscoding probability r in the robustification procedure. A large r may underestimate the activation clusters, while a small r tends to allow more false positives. The choice of r is often subjective. One may use information from, say, previous studies, to find an appropriate r in order to balance this trade-off. If no prior information is available, [Choudhuri, Ghosal and Roy \(2007\)](#) proposed letting r be a small number between 0.01 and 0.1. In practice, we suggest experimenting with several r values and choosing the one that gives the best results.

APPENDIX: MCMC ALGORITHMS FOR POSTERIOR INFERENCE

A.1. Probit link. Let $\mathbf{y} = (y_1, \dots, y_n)^T$ be the random vector of binary observations measured and $\mathbf{x} = (x_1, \dots, x_n)^T$ the corresponding covariate values, where each x_j has one or two component variables. Let $\mathbf{w} = (w_1, \dots, w_n)^T$ be some unobservable latent variable. Following [Holmes and Held \(2006\)](#), the probit model can be written as

$$(6) \quad \begin{aligned} y_i &= \begin{cases} 1, & \text{if } w_i > 0, \\ 0, & \text{if } w_i \leq 0, \end{cases} \\ w_i &= z_i + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1), \end{aligned}$$

where $\mathbf{z} = (z_1, \dots, z_n)^T$ is the adaptive GMRF described in Section 2.1. Since y_i are now deterministic conditional on the sign of the w_i , we have $P(y_i = 0|z_i) = P(w_i \leq 0|z_i) = \Phi(-z_i)$, where Φ is the standard Gaussian c.d.f.

As mentioned earlier, the half- t prior on δ can be written as $\delta \stackrel{D}{=} |\xi|\theta$, where $\xi \sim N(0, 1)$ and $\theta^2 \sim \text{IG}(\rho/2, \rho S^2/2)$. A redundant multiplicative reparameterization can be applied to model (6):

$$\begin{aligned} y_i &= \begin{cases} 1, & \text{if } w_i > 0, \\ 0, & \text{if } w_i \leq 0, \end{cases} \\ w_i &= \xi \eta_i + \varepsilon_i, \quad \varepsilon_i \stackrel{\text{i.i.d.}}{\sim} N(0, 1), \end{aligned}$$

where $\boldsymbol{\eta} = (\eta_1, \dots, \eta_n)^T$ has a GMRF prior density

$$[\boldsymbol{\eta}|\theta^2, \boldsymbol{\gamma}] \propto |\theta^{-2} \mathbf{A}_\gamma|_+^{1/2} \exp\left(-\frac{1}{2\theta^2} \boldsymbol{\eta}^T \mathbf{A}_\gamma \boldsymbol{\eta}\right),$$

with $\mathbf{A}_\gamma = \mathbf{B}'_m \text{diag}(\boldsymbol{\gamma}) \mathbf{B}_m$ for $m = 1, 2$. This expanded model form allows conditionally conjugate prior distributions for both ξ and θ , and these parameters are independent in the conditional posterior distribution [Gelman (2006); Gelman et al. (2008)]. Letting d be the dimension of the null space of \mathbf{A}_γ , the full conditional distributions are listed below:

- $(\boldsymbol{\eta}|\theta^2, \xi, \boldsymbol{\gamma}, \mathbf{w}) \sim N_n(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta)$, where $\boldsymbol{\mu}_\eta = \xi \boldsymbol{\Sigma}_\eta \mathbf{w}$ and $\boldsymbol{\Sigma}_\eta = (\xi^2 \mathbf{I}_n + \mathbf{A}_\gamma/\theta^2)^{-1}$;
- $(\xi|\boldsymbol{\eta}, \mathbf{w}) \sim N(\mu_\xi, \sigma_\xi^2)$, where $\mu_\xi = \sigma_\xi^2 \boldsymbol{\eta}' \mathbf{w}$ and $\sigma_\xi^2 = (1 + \boldsymbol{\eta}' \boldsymbol{\eta})^{-1}$;
- $(w_i|\xi, \boldsymbol{\eta}, \mathbf{y}) \sim \begin{cases} N(\xi \eta_i, 1)I(w_i > 0), & \text{if } y_i = 1, \\ N(\xi \eta_i, 1)I(w_i \leq 0), & \text{if } y_i = 0; \end{cases}$
- $(\gamma_j|\theta^2, \boldsymbol{\eta}) \sim \text{IG}(\frac{\nu+1}{2}, \frac{1}{2\theta^2} \tilde{\eta}_j^2 + \frac{1}{2})$, where $\tilde{\boldsymbol{\eta}} = \mathbf{B}_m \boldsymbol{\eta}$ ($m = 1, 2$);
- $(\theta^2|\boldsymbol{\eta}, \boldsymbol{\gamma}) \sim \text{IG}(\frac{n-d}{2} + \frac{\rho}{2}, \frac{1}{2} \boldsymbol{\eta}' \mathbf{A}_\gamma \boldsymbol{\eta} + \frac{\rho S^2}{2})$.

Note that $\boldsymbol{\Sigma}_\eta$ is a banded matrix and we can thus use the banded Cholesky decomposition to simulate $\boldsymbol{\eta}$ with the cost of $\mathcal{O}(n)$. The quantities w_i have independent truncated normal distributions and are also straightforward to sample from.

A.2. Logit link. Again, we use data augmentation and overparameterization to write the logistic regression model as

$$(7) \quad \begin{aligned} y_i &= \begin{cases} 1, & \text{if } w_i > 0, \\ 0, & \text{if } w_i \leq 0, \end{cases} \\ w_i &= \xi \eta_i + \varepsilon_i, \quad \varepsilon_i \sim N(0, \lambda_i), \\ \lambda_i &= (2\kappa_i)^2, \quad \kappa_i \sim \text{KS}, \end{aligned}$$

where KS denotes the Kolmogorov–Smirnov distribution [e.g., Devroye (1986)]. In this case, ε_i has the form of a scale mixture of normals with a marginal logistic distribution.

To improve mixing of the Markov chains, we update $\{\mathbf{w}, \boldsymbol{\lambda}\}$ jointly given $\{\xi, \boldsymbol{\eta}\}$,

$$[\mathbf{w}, \boldsymbol{\lambda}|\xi, \boldsymbol{\eta}, \mathbf{y}] = [\mathbf{w}|\xi, \boldsymbol{\eta}, \mathbf{y}][\boldsymbol{\lambda}|\mathbf{w}, \xi, \boldsymbol{\eta}].$$

Letting $\boldsymbol{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_n)$, the posterior conditional distributions are as follows:

- $(\boldsymbol{\eta}|\theta^2, \xi, \boldsymbol{\gamma}, \mathbf{w}, \boldsymbol{\lambda}) \sim N_n(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta)$, where $\boldsymbol{\mu}_\eta = \xi \boldsymbol{\Sigma}_\eta \boldsymbol{\Lambda} \mathbf{w}$ and $\boldsymbol{\Sigma}_z = (\xi^2 \boldsymbol{\Lambda} + \mathbf{A}_\gamma/\theta^2)^{-1}$;
- $(\xi|\boldsymbol{\eta}, \mathbf{w}, \boldsymbol{\lambda}) \sim N(\mu_\xi, \sigma_\xi^2)$, where $\mu_\xi = \sigma_\xi^2 \boldsymbol{\eta}' \boldsymbol{\Lambda} \mathbf{w}$ and $\sigma_\xi^2 = (1 + \boldsymbol{\eta}' \boldsymbol{\Lambda} \boldsymbol{\eta})^{-1}$;
- $(w_i|\xi, \boldsymbol{\eta}, \mathbf{y}) \sim \begin{cases} \text{Logistic}(\xi \eta_i, 1)I(w_i > 0), & \text{if } y_i = 1, \\ \text{Logistic}(\xi \eta_i, 1)I(w_i \leq 0), & \text{if } y_i = 0; \end{cases}$
- $[\lambda_i|w_i, \xi, \eta_i] \propto \lambda_i^{-1} \exp\{-\frac{1}{2\lambda_i} (w_i - \xi \eta_i)^2\} \text{KS}(\frac{\sqrt{\lambda_i}}{2})$;
- $(\gamma_j|\theta^2, \boldsymbol{\eta}) \sim \text{IG}(\frac{\nu+1}{2}, \frac{1}{2\theta^2} \tilde{\eta}_j^2 + \frac{\nu}{2})$, where $\tilde{\boldsymbol{\eta}} = \mathbf{B}_m \boldsymbol{\eta}$ ($m = 1, 2$);
- $(\theta^2|\boldsymbol{\eta}, \boldsymbol{\gamma}) \sim \text{IG}(\frac{n-d}{2} + \frac{\rho}{2}, \frac{1}{2} \boldsymbol{\eta}' \mathbf{A}_\gamma \boldsymbol{\eta} + \frac{\rho S^2}{2})$.

The Logistic(α, β) denotes the density function of the logistic distribution with mean α and scale parameter β [Devroye (1986), page 39]. Sampling from the truncated logistic distribution can be done efficiently by the inversion method. Although it is not a standard task, sampling λ_i is simple using a rejection method as outlined in Holmes and Held (2006).

A.3. Other scale mixtures of normal links. The auxiliary variable sampling scheme described above can easily be generalized to work for any link function H that can be represented as scale mixtures of normal cdfs, and, hence,

$$H(t) = \int_0^\infty \Phi\left(\frac{t}{\sqrt{v}}\right) dG(v),$$

where v follows some continuous or discrete distribution G on $(0, \infty)$. A wide class of continuous, unimodal and symmetric distributions on the real line may be constructed as scale mixtures of normals. Many examples, such as discrete mixtures or contaminated normals, the Student t family, logistic, Laplace or double-exponential, and the stable family, are well known; see, for example, Andrews and Mallows (1974).

Similarly, we introduce two sets of latent variables $\mathbf{w} = (w_1, \dots, w_n)^T$ and $\mathbf{v} = (v_1, \dots, v_n)^T$ such that $(w_i | \mathbf{z}, \mathbf{v}) \sim N(z_i, v_i)$, $v_i \stackrel{\text{i.i.d.}}{\sim} G$, and $y_i = I(w_i > 0)$. Then, conditional on \mathbf{z} , the y_i 's are independent Bernoulli random variables with success probability $H(z_i)$. Suppose G has a Lebesgue density or probability mass function g . Let $z_i = \xi \eta_i$ and $\mathbf{V} = \text{diag}(v_1, \dots, v_n)$. Then, the posterior conditional distributions are as follows:

- $(\eta | \theta^2, \xi, \boldsymbol{\gamma}, \mathbf{w}, \mathbf{v}) \sim N_n(\boldsymbol{\mu}_\eta, \boldsymbol{\Sigma}_\eta)$, where $\boldsymbol{\mu}_\eta = \xi \mathbf{V} \boldsymbol{\Sigma}_\eta \boldsymbol{\gamma}$ and $\boldsymbol{\Sigma}_\eta = (\xi^2 \mathbf{V} + \mathbf{A}_\gamma / \theta^2)^{-1}$;
- $(\xi | \boldsymbol{\eta}, \mathbf{w}, \mathbf{v}) \sim N(\mu_\xi, \sigma_\xi^2)$, where $\mu_\xi = \sigma_\xi^2 \boldsymbol{\eta}' \mathbf{V} \mathbf{w}$ and $\sigma_\xi^2 = (1 + \boldsymbol{\eta}' \mathbf{V} \boldsymbol{\eta})^{-1}$;
- $(w_i | \xi, \boldsymbol{\eta}, \mathbf{v}, \mathbf{y}) \sim \begin{cases} N(\xi \eta_i, v_i) I(w_i > 0), & \text{if } y_i = 1, \\ N(\xi \eta_i, v_i) I(w_i \leq 0), & \text{if } y_i = 0; \end{cases}$
- $[v_i | \xi, w_i, \eta_i] \propto v_i^{-1/2} \exp\{-\frac{1}{2v_i}(w_i - \xi \eta_i)^2\} g(v_i)$;
- $(\gamma_j | \theta^2, \boldsymbol{\eta}) \sim \text{IG}(\frac{\nu+1}{2}, \frac{1}{2\theta^2} \tilde{\eta}_j^2 + \frac{\nu}{2})$, where $\tilde{\boldsymbol{\eta}} = \mathbf{B}_m \boldsymbol{\eta}$ ($m = 1, 2$);
- $(\theta^2 | \boldsymbol{\eta}, \boldsymbol{\gamma}) \sim \text{IG}(\frac{n-d}{2} + \frac{\rho}{2}, \frac{1}{2} \boldsymbol{\eta}' \mathbf{A}_\gamma \boldsymbol{\eta} + \frac{\rho S^2}{2})$.

Thus, a Gibbs sampler can be used to sample joint posterior distributions. The only difficult part is sampling θ_i . For a Student t link, the mixing distribution G is an inverse gamma distribution, as is the full conditional of each v_i . For the Laplace link, the G is an exponential distribution and the v_i^{-1} follows an inverse Gaussian conditional distribution. Therefore, one can directly sample v_i 's for those two links. If $[v_i | \xi, w_i, \eta_i]$ does not correspond to any regular density, the samples may be drawn via acceptance-rejection sampling.

Acknowledgment. The authors thank Tor Wager for the meta-analysis data. We would like to thank the anonymous referees and Associate Editor for valuable comments that led to considerable improvement upon the original version of this paper. We also thank Tor Wager for the meta-analysis data.

REFERENCES

- ALBERT, J. H. and CHIB, S. (1993). Bayesian analysis of binary and polychotomous response data. *J. Amer. Statist. Assoc.* **88** 669–679. [MR1224394](#)
- ANDREWS, D. F. and MALLOWS, C. L. (1974). Scale mixtures of normal distributions. *J. Roy. Statist. Soc. Ser. B* **36** 99–102. [MR0359122](#)
- BREZGER, A., FAHRMEIR, L. and HENNERFEIND, A. (2007). Adaptive Gaussian Markov random fields with applications in human brain mapping. *J. Roy. Statist. Soc. Ser. C* **56** 327–345. [MR2370993](#)
- CARTER, C. K. and KOHN, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models. *Biometrika* **83** 589–601. [MR1423876](#)
- CARVALHO, C. M., POLSON, N. G. and SCOTT, J. G. (2010). The horseshoe estimator for sparse signals. *Biometrika* **97** 465–480. [MR2650751](#)
- CHOUDHURI, N., GHOSAL, S. and ROY, A. (2007). Nonparametric binary regression using a Gaussian process prior. *Stat. Methodol.* **4** 227–243. [MR2368147](#)
- CRAINICEANU, C. M., RUPPERT, D., CARROLL, R. J., ADARSH, J. and GOODNER, B. (2007). Spatially adaptive penalized splines with heteroscedastic errors. *J. Comput. Graph. Statist.* **16** 265–288.
- DALE, A. M., FISCHL, B. and SERENO, M. I. (1999). Cortical surface-based analysis. I. Segmentation and surface reconstruction. *Neuroimage* **9** 179–194.
- DEVROYE, L. (1986). *Non-Uniform Random Variate Generation*. Springer, New York.
- DEY, D. K., GHOSH, S. K. and MALLICK, B. K., eds. (2000). *Generalized Linear Models: A Bayesian Perspective*. *Biostatistics* **5**. Dekker, New York. [MR1893779](#)
- FISCHL, B., SERENO, M. I. and DALE, A. M. (1999). Cortical surface-based analysis. II: Inflation, flattening, and a surface-based coordinate system. *Neuroimage* **9** 195–207.
- GELMAN, A. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by Browne and Draper). *Bayesian Anal.* **1** 515–533 (electronic). [MR2221284](#)
- GELMAN, A., VAN DYK, D. A., HUANG, Z. and BOSCARDIN, W. J. (2008). Using redundant parameterizations to fit hierarchical models. *J. Comput. Graph. Statist.* **17** 95–122. [MR2424797](#)
- GU, C. (1990). Adaptive spline smoothing in non-Gaussian regression models. *J. Amer. Statist. Assoc.* **85** 801–807. [MR1138360](#)
- HASTIE, T. J. and TIBSHIRANI, R. J. (1990). *Generalized Additive Models*. *Monographs on Statistics and Applied Probability* **43**. Chapman & Hall, London. [MR1082147](#)
- HOLMES, C. C. and HELD, L. (2006). Bayesian auxiliary variable models for binary and multinomial regression. *Bayesian Anal.* **1** 145–168 (electronic). [MR2227368](#)
- HOLMES, C. C. and MALLICK, B. K. (2003). Generalized nonlinear modeling with multivariate free-knot regression splines. *J. Amer. Statist. Assoc.* **98** 352–368. [MR1995711](#)
- KANG, J., JOHNSON, T. D., NICHOLS, T. E. and WAGER, T. D. (2011). Meta analysis of functional neuroimaging data via Bayesian spatial point processes. *J. Amer. Statist. Assoc.* **106** 124–134. [MR2816707](#)
- KOBER, H., BARRETT, L. F., JOSEPH, J., BLISS-MOREAU, E., LINDQUIST, K. and WAGER, T. D. (2008). Functional grouping and cortical-subcortical interactions in emotion: A meta-analysis of neuroimaging studies. *Neuroimage* **42** 998–1031.
- KRIVOBOKOVA, T., CRAINICEANU, C. M. and KAUEMANN, G. (2008). Fast adaptive penalized splines. *J. Comput. Graph. Statist.* **17** 1–20. [MR2424792](#)

- LINDQUIST, M. A. (2008). The statistical analysis of fMRI data. *Statist. Sci.* **23** 439–464. [MR2530545](#)
- LOADER, C. (1999). *Local Regression and Likelihood*. Springer, New York. [MR1704236](#)
- MCCULLAGH, P. and NELDER, J. (1989). *Generalized Linear Models*, 2nd ed. Chapman & Hall/CRC, Boca Raton, FL.
- O’SULLIVAN, F., YANDELL, B. S. and RAYNOR, W. J. JR. (1986). Automatic smoothing of regression functions in generalized linear models. *J. Amer. Statist. Assoc.* **81** 96–103. [MR0830570](#)
- PENNY, W. D., TRUJILLO-BARRETO, N. J. and FRISTON, K. J. (2005). Bayesian fMRI time series analysis with spatial priors. *Neuroimage* **24** 350–362.
- PSARAKIS, S. and PANARETOS, J. (1990). The folded t distribution. *Comm. Statist. Theory Methods* **19** 2717–2734. [MR1086030](#)
- RUE, H. and HELD, L. (2005). *Gaussian Markov Random Fields: Theory and Applications. Monographs on Statistics and Applied Probability* **104**. Chapman & Hall/CRC, Boca Raton, FL. [MR2130347](#)
- RUE, H., MARTINO, S. and CHOPIN, N. (2009). Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **71** 319–392. [MR2649602](#)
- RUPPERT, D. and CARROLL, R. J. (2000). Spatially-adaptive penalties for spline fitting. *Aust. N. Z. J. Stat.* **42** 205–223.
- SPIEGELHALTER, D. J., BEST, N. G., CARLIN, B. P. and VAN DER LINDE, A. (2002). Bayesian measures of model complexity and fit. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **64** 583–639. [MR1979380](#)
- TALAIRACH, J. and TOURNOUX, P. (1988). *Co-planar Stereotaxic Atlas of the Human Brain: 3-Dimensional Proportional System—an Approach to Cerebral Imaging*. Thieme Medical Publishers, New York.
- TIPPING, M. E. (2001). Sparse Bayesian learning and the relevance vector machine. *J. Mach. Learn. Res.* **1** 211–244. [MR1875838](#)
- TRIPPA, L. and MULIERE, P. (2009). Bayesian nonparametric binary regression via random tessellations. *Statist. Probab. Lett.* **79** 2273–2280. [MR2591984](#)
- TURKELTAUB, P., EDEN, G., JONES, K. and ZEFFIRO, T. A. T. (2002). Meta-analysis of the functional neuroanatomy of single-word reading: Method and validation. *NeuroImage* **16** 765–780.
- WAGER, T. D., JONIDES, J. and READING, S. (2004). Neuroimaging studies of shifting attention: A meta-analysis. *Neuroimage* **22** 1679–1693.
- WAGER, T. D., LINDQUIST, M. A. and KAPLAN, L. (2007). Meta-analysis of functional neuroimaging data: Current and future directions. *Social Cognitive and Affective Neuroscience* **2** 150–158.
- WAGER, T. D., BARRETT, L. F., BLISS-MOREAU, E., LINDQUIST, K., DUNCAN, S., KOBER, H., JOSEPH, J., DAVIDSON, M. and MIZE, J. (2008). The neuroimaging of emotion. In *Handbook of Emotion* (M. Lewis, ed.) 249–271. Guilford Press, New York.
- WAGER, T. D., LINDQUIST, M. A., NICHOLS, T. E., KOBER, H. and VAN SNELLENBERG, J. X. (2009). Evaluating the consistency and specificity of neuroimaging data using meta-analysis. *Neuroimage* **45** S210–S221.
- WAHBA, G., WANG, Y., GU, C., KLEIN, R. and KLEIN, B. (1995). Smoothing spline ANOVA for exponential families, with application to the Wisconsin Epidemiological Study of Diabetic Retinopathy. *Ann. Statist.* **23** 1865–1895. [MR1389856](#)
- WIPER, M. P., GIRÓN, F. J. and PEWSEY, A. (2008). Objective Bayesian inference for the half-normal and half- t distributions. *Comm. Statist. Theory Methods* **37** 3165–3185. [MR2467759](#)
- WOOD, S. A. and KOHN, R. (1998). A Bayesian approach to robust binary nonparametric regression. *J. Amer. Statist. Assoc.* **93** 203–213.
- WOOD, S. A., KOHN, R., COTTET, R., JIANG, W. and TANNER, M. (2008). Locally adaptive nonparametric binary regression. *J. Comput. Graph. Statist.* **17** 352–372. [MR2439964](#)

- YUE, Y., LOH, J. M. and LINDQUIST, M. A. (2010). Adaptive spatial smoothing of fMRI images. *Stat. Interface* **3** 3–13. [MR2609707](#)
- YUE, Y. R. and LOH, J. M. (2011). Bayesian semiparametric intensity estimation for inhomogeneous spatial point processes. *Biometrics* **67** 937–946.
- YUE, Y. and SPECKMAN, P. L. (2010). Nonstationary spatial Gaussian Markov random fields. *J. Comput. Graph. Statist.* **19** 96–116. [MR2654402](#)

Y. R. YUE
DEPARTMENT OF STATISTICS AND CIS
ZICKLIN SCHOOL OF BUSINESS
CITY UNIVERSITY OF NEW YORK
ONE BERNARD BARUCH WAY
NEW YORK, NEW YORK 10010
USA
E-MAIL: yu.yue@baruch.cuny.edu

M. A. LINDQUIST
DEPARTMENT OF STATISTICS
COLUMBIA UNIVERSITY
1255 AMSTERDAM AVE
NEW YORK, NEW YORK 10027
USA
E-MAIL: martin@stat.columbia.edu

J. M. LOH
AT&T LABS-RESEARCH
180 PARK AVE-BUILDING 103
FLORHAM PARK, NEW JERSEY 07932
USA
E-MAIL: loh@research.att.com