Bayesian Analysis (2015) **TBA**, Number TBA, pp. 1–27

Pre-surgical fMRI Data Analysis Using a Spatially Adaptive Conditionally Autoregressive Model

Zhuqing Liu[*], Veronica J. Berrocal[†], Andreas J. Bartsch[‡,§,¶], and Timothy D. Johnson[‖]

Abstract. Spatial smoothing is an essential step in the analysis of functional magnetic resonance imaging (fMRI) data. One standard smoothing method is to convolve the image data with a three-dimensional Gaussian kernel that applies a fixed amount of smoothing to the entire image. In pre-surgical brain image analysis where spatial accuracy is paramount, this method, however, is not reasonable as it can blur the boundaries between activated and deactivated regions of the brain. Moreover, while in a standard fMRI analysis strict false positive control is desired, for pre-surgical planning false negatives are of greater concern. To this end, we propose a novel spatially adaptive conditionally autoregressive model with variances in the full conditional of the means that are proportional to error variances, allowing the degree of smoothing to vary across the brain. Additionally, we present a new loss function that allows for the asymmetric treatment of false positives and false negatives. We compare our proposed model with two existing spatially adaptive conditionally autoregressive models. Simulation studies show that our model outperforms these other models; as a real model application, we apply the proposed model to the pre-surgical fMRI data of two patients to assess peri- and intra-tumoral brain activity.

Keywords: fMRI analysis, spatially adaptive CAR models, loss function, pre-surgical mapping.

1 Introduction

Over the past two decades, researchers in the cognitive neurosciences have used functional magnetic resonance imaging (fMRI) to study changes in brain activation brought on by various experimental stimuli (Ogawa et al., 1990; Kwong et al., 1992). More recently researchers have found fMRI useful in studying functional deficits in patients

[*]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, zhuqingl@umich.edu

[†]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, berrocal@umich.edu

[‡]Department of Neuroradiology, University of Heidelberg, Im Neuenheimer Feld 400, 69120 Heidelberg, Germany, bartsch@radvisory.net

[§]Department of Neuroradiology, University of Wuerzburg, Joseph-Schneider-Str. 11, 97080 Wuerzburg, Germany

[¶]FMRIB Centre, Department of Clinical Neurology, University of Oxford, Oxford, UK

[‖]Department of Biostatistics, School of Public Health, University of Michigan, Ann Arbor, MI 48109, tdjtdj@umich.edu

 DOI: 10.1214/15-BA972

with neurological diseases or psychiatric disorders, as well as connectivity patterns during rest. In addition, there is growing interest in using fMRI to map out functionally relevant brain regions as a pre-surgical tool to aid neurosurgeons during the planning stages of tumor resection (Bartsch et al., 2006; Sunaert, 2006; Bookheimer, 2007; Durnez et al., 2013). Brain tumor resection is an arduous task. The surgeon's goal is to remove as much tumor as possible while preserving as much healthy brain tissue as possible; especially avoiding the removal, and damage, of healthy regions of the brain that are vital to the patient's quality of life. The neurosurgeon may try to achieve this goal by "awake craniotomy" and electrical stimulation mapping (ESM). That is, once the patient is conscious, the surgeon begins mapping out those regions of the brain near or within the tumor, i.e., peri- or intra-tumor, that are vital to the patient's quality of life by asking the patient to perform certain tasks. During these tasks, the surgeon probes the brain with an electrode that delivers a small amount of electrical current. Once the task is disrupted, the surgeon places a small marker on that part of the brain that is responsible for the task and will avoid that region during surgery. This surgical mapping and tumor resection may take up to twelve hours or longer depending on the location of the tumor relative to vital brain circuits and regions. These lengthy procedures place great stress on both the patient and the surgeon. Additionally, ESM is subject to both false-negative and false-positive results and is complicated by seizures in up to 15% of the patients (Nossek et al., 2013). Mapping out these vital regions via fMRI prior to surgery allows the surgeon to quickly hone in on these regions for verification, thus speeding up the ESM and tumor resection. Thus, spatial precision of functionally relevant regions, as defined by fMRI scans, is of utmost importance and is the motivation for the work presented in this manuscript.

After acquisition of fMRI data, several steps are typically performed to map brain activity since fMRI data, time series of three-dimensional volume data, each consisting of thousands of uniformly spaced volume elements, called 'voxels' (Lindquist, 2008), are intrinsically noisy. The steps typically are: (i) Pre-processing. This includes correction for motion and geometric distortions, image registration, temporal filtering, spatial smoothing and pre-whitening; (ii) Fitting a general linear model to every time series at each voxel, independently of one another. This step is referred to as a massive univariate approach. The resulting image of standardized parameter values is called a Z-statistic image, or map. In reality, the resulting standardized parameters all follow a student-T distribution, however, since there are typically hundreds to thousands of observations in each time series at each voxel, the normal distribution is a very good approximation; and (iii) Performing hypothesis testing with correction for multiple comparisons. The two most commonly used analysis tools are the Statistical Parametric Mapping (SPM) (Friston et al., 1994) software and the FMRIB Software Library (FSL) (Smith et al., 2004; Woolrich et al., 2009). In the pre-processing step, the fMRI data are spatially smoothed by convolving the data with a Gaussian kernel with spatial extent described by the full width of the kernel at half its maximum height (FWHM) (Friston et al., 1995). Spatial smoothing is an important preprocessing step that is performed to achieve two goals. First, it increases the signal-to-noise ratio. Second, and more importantly, it helps guarantee that the assumptions of Gaussian random field theory, a method commonly used in the following multiple comparisons procedure (Worsley et al., 1996), are satisfied.

Although the Gaussian kernel approach is the most popular approach for smoothing fMRI images, Holmes et al. (1996) point out several associated problems, one of which is of great concern for pre-surgical planning: the spatial smoothing pre-processing step uses a Gaussian kernel with a fixed FWHM and does not locally adapt the amount of smoothing based on the data. This can cause some regions to be under-smoothed while others to be over-smoothed, blurring out the boundaries of activated or deactivated regions with null regions. (Null regions or voxels are defined as regions or voxels where the Blood-Oxygenation-Level-Dependent (BOLD) signal (Ogawa et al., 1992) does not change significantly between the fMRI control condition and the fMRI experimental condition. Activated regions or voxels are those regions or voxels where the BOLD signal is significantly larger during the experimental condition as compared to the control condition. Deactivated regions or voxels are those regions or voxels where the BOLD signal is significantly smaller during the experimental condition compared to the control condition. A false negative occurs when a truly activated or deactivated voxel is classified as null while a false positive occurs when a truly null voxel is classified as either activated or deactivated.) Furthermore, for a small (relative to the FWHM of the smoothing kernel) activated or deactivated region, global smoothing can completely obliterate its significance. Thus, one important challenge facing pre-surgical fMRI is to avoid blurring any sharp boundaries of activated or deactivated regions and to avoid false-negatives by over-smoothing, ensuring spatial accuracy (Yoo et al., 2004).

Our goal in this paper is to develop a statistical model that will adaptively smooth pre-surgical fMRI data; smoothing more in regions where smoothness is warranted and smoothing less in regions where it is not, for example, at the interface between activated and null regions. We fit our model to the Z-statistic image in native space. (That is, in the geometry of the subject's brain. We do not register to a standard template, however, geometric distortions are corrected, if necessary.) We use the FSL software to perform both the preprocessing step and the massive univariate analysis that results in a Z-statistic image. However, it is important to note that we do not spatially smooth the data in the preprocessing step. A classical statistical method to smooth images is to fit a linear model to the data with spatial random effects using an autonormal model (Besag, 1974). Typically, the spatial random effects are assigned a conditionally autoregressive (CAR) prior (Besag et al., 1991, BYM). In the BYM model, smoothing is controlled by a global smoothing parameter, the spatial variance, and thus a fixed amount of smoothing is applied to the entire image. However, global smoothing is not always warranted and certain parameters in the CAR model can be assigned hyperprior distributions that allow adaptive smoothing. For examples, see Brewer and Nolan (2007, BN) and Reich and Hodges (2008, RH).

Brewer and Nolan (2007) allow for adaptive local smoothing by assigning each site a variance parameter and setting the parameters in the autonormal model in (1) to yield (7), while Reich and Hodges (2008) set them so to yield (8). The two models differ on how the spatially adaptive variances of the full conditionals of the means are defined, see (7) and (8). Though the RH and BN models are designed to achieve adaptive local smoothing, application of these models to our fMRI data proved problematic. The BN model over-smoothed the Z-statistic image while the RH model failed to converge depending on the initial state of the Markov chain. These problems are most likely due to

the size of the fMRI data (approximately 47000 voxels). We discuss these issues further in Sections 4 and 5. More recently, Yue et al. (2010) introduced an adaptive spatial smoothing model that uses a non-stationary spatial Gaussian Markov random field and applied it to fMRI data. However, they only consider single slices—two dimensions in space of the 3-dimensional fMRI time series data—applying their model to every slice in the time series (computational considerations prevented them from modeling the full 3-dimensional time series data).

Another approach to avoid blurring of boundaries is to use a Potts model (Potts, 1952). Woolrich et al. (2005) proposed a parametric Potts model for image segmentation that could be used for pre-surgical fMRI. Johnson et al. (2013) proposed a Bayesian non-parametric Potts model and showed that the non-parametric Potts model has better statistical properties if the parametric Potts model is mis-specified, and performed on par with the parametric model when it is correctly specified. While the non-parametric Potts model shows good statistical properties, there are practical issues with computational cost. Johnson et al. (2013) estimate two model parameters in the potential function and both depend on an intractable normalizing constant. Thus, they use a path sampling (Gelman and Meng, 1998) approach to estimate the log-ratio of these normalizing constants on a two-dimensional grid of parameter values. The log-ratios are determined via simulation in a pre-processing step and this step can take one week or longer. Since pre-surgical fMRI analyses are performed in native space, this pre-processing step must be performed for each individual. A possible delay in surgery to obtain the final analysis is not acceptable. Recently, we learned of a method (Murray et al., 2006) that can estimate these parameters in a Potts model without the pre-processing step. However, we did not find any advantage with respect to computation time over the method proposed by Johnson et al. (2013). Therefore, we propose a novel model that adaptively and locally smooths the fMRI data. Compared to the BN and RH models, our model offers a more intuitive interpretation and presents less problems during model fitting. In particular, in our model we let the variances in the full conditional of the means be proportional to the error variances, allowing the degree of smoothing to vary across the brain. We call our model a conditionally weighted adaptive smoothing model (CWAS; see Section 2.2.1 for details).

In pre-surgical fMRI data, false negatives (i.e., voxels falsely classified as null) are more dangerous to the patient than false positives (i.e., voxels falsely classified as activated or de-activated) since the former may result in the surgeon damaging healthy tissue vital to quality of life (Haller and Bartsch, 2009). Building upon the work of Müller et al. (2007), we propose a loss function to identify functionally active regions in the brain that asymmetrically penalizes false negatives and false positives. We use the proposed loss function to compare the performance of our model to that of the BN and RH models. Simulation studies show that the proposed model outperforms the BN and RH models in terms of false negative rates.

The remainder of this paper is organized as follows. We begin in Section 2 with a description of the proposed CWAS model, a brief overview of the BN and RH models, and our proposed loss function. We then discuss model implementation details in Section 3. In Section 4, we present results from our simulation studies. We then apply our

model to two pre-surgical fMRI data sets in Section 5 and compare results with those from the BN and RH models. We conclude the paper in Section 6 with a discussion of the contribution of this paper and future directions.

2 Model

2.1 The CAR model

Consider an unsmoothed 3-dimensional Z-statistic image. We use $\boldsymbol{Y}$ to represent its Z-statistic intensity for the set of N voxels, where Y_i is the intensity of the ith voxel, for $i = 1, \ldots, N$. We assume a first order neighborhood system with two voxels, i and i', neighbors if they share a common face, and we denote it by $i \sim i'$. Suppose μ_i and ϵ_i are the mean intensity and the random measurement error of voxel i, respectively, with error $\epsilon_i \sim \mathrm{N}(0, \sigma_i^2)$, where $\mathrm{N}(a, b)$ denotes the normal distribution with mean a and variance b. Denote the mean intensity vector by $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)^T$. In this context, $\boldsymbol{Y}$ is the Z-statistic image and $\boldsymbol{\mu}$ represents the smoothed Z-statistic image. We assume $Y_i = \mu_i + \epsilon_i$, i.e., $Y_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$.

A classical statistical model to smooth images is the CAR model (Besag et al., 1991). The CAR model is a specific form of the auto-normal model (Besag, 1974), the latter being defined by the set of full conditionals:

$$[\mu_i \mid \boldsymbol{\mu}_{(-i)}, \boldsymbol{B}, \phi_i^2] \sim \mathrm{N}\left(\sum\nolimits_j b_{ij}\mu_j,\ \phi_i^2\right) \tag{1}$$

where $\boldsymbol{\mu}_{(-i)}$ is the $(n-1)$-dimensional vector obtained from $\boldsymbol{\mu}$ by removing μ_i, and $\boldsymbol{B}$ is an $N \times N$ matrix with elements b_{ij}, where $b_{ij} = 0$ if $i = j$.

Set $w_{ij} = 1$ if and only if voxels i and j are neighbors (note: a voxel is not a neighbor of itself), set $w_{i+} = \sum_{j=1}^{N} w_{ij}$, and let $\boldsymbol{W}$ be the $N \times N$ matrix with elements w_{ij}. The CAR model is obtained when $b_{ij} = w_{ij}/w_{i+}$ and $\phi_i^2 = \tau^2/w_{i+}$ with $\tau \in \mathbb{R}^+$. Then, the full conditionals in (1) are

$$[\mu_i \mid \boldsymbol{\mu}_{(-i)}, \boldsymbol{W}, \tau^2] \sim \mathrm{N}\left(\sum\nolimits_j (w_{ij}/w_{i+})\,\mu_j,\ \tau^2/w_{i+}\right). \tag{2}$$

Here $\sum_j (w_{ij}/w_{i+})\,\mu_j$ indicates the average of the μ_js at voxels that are neighbors to voxel i and thus the resulting μ_i is smoothed towards the mean of its neighbors. The amount of smoothing in (2) is controlled by a global parameter τ^2 while in the original auto-normal model the amount of smoothing is controlled by the voxel-specific variance ϕ_i^2.

By Brook's Lemma (Brook, 1964), the full conditionals in (2) imply the following joint prior on $\boldsymbol{\mu}$ up to a normalizing constant:

$$\pi(\boldsymbol{\mu} \mid \tau^2) \propto \exp\left\{-0.5\tau^{-2} \sum\nolimits_{i \sim j} (\mu_i - \mu_j)^2\right\},$$

which is also called pairwise difference prior since it only depends on the differences of all neighbor pairs (Besag, 1993).

2.2 Spatially adaptive CAR models

The CWAS model

We define our spatially adaptive CAR model on $\boldsymbol{\mu}$ by specifying a specific form for the variance ϕ_i^2 in (1). Instead of using a global parameter τ^2 for all voxels as in the CAR model, we let ϕ_i^2 vary across the brain and model it to be proportional to the error variance σ_i^2, i.e., $\phi_i^2 = c_i\sigma_i^2$, where $c_i > 0$. We still assume $b_{ij} = w_{ij}/w_{i+}$. Then, our spatially adaptive CAR model is defined through the following conditional distributions:

$$[\mu_i \mid \boldsymbol{\mu}_{(-i)}, \boldsymbol{W}, c_i, \sigma_i^2] \sim \mathrm{N}\left(\sum\nolimits_j (w_{ij}/w_{i+})\,\mu_j,\ c_i\sigma_i^2\right),\ i = 1, \ldots, N. \tag{3}$$

To allow for more flexibility, we let the error variances vary spatially as well by placing a CAR prior on their logarithm, that is,

$$[\ln(\sigma_i^2) \mid \ln(\boldsymbol{\sigma}_{(-i)}^2), \boldsymbol{W}, \lambda^2] \sim \mathrm{N}\left(\sum\nolimits_j (w_{ij}/w_{i+})\ln(\sigma_j^2),\ \lambda^2/w_{i+}\right). \tag{4}$$

Equations (6) and (4) are the motivation for calling our model the conditionally weighted adaptive smoothing (CWAS) model.

The entire CWAS model is as follows:

$$\begin{aligned}
[y_i \mid \mu_i, \sigma_i^2] &\overset{ind}{\sim} \mathrm{N}\left(\mu_i,\ \sigma_i^2\right),\ i = 1, \ldots, N,\\
\left[\mu_i \mid \boldsymbol{\mu}_{(-i)}, c_i, \sigma_i^2\right] &\sim \mathrm{N}\left(\sum\nolimits_j (w_{ij}/w_{i+})\,\mu_j,\ c_i\sigma_i^2\right), i = 1, \ldots, N,\\
\left[\ln(\sigma_i^2) \mid \ln(\boldsymbol{\sigma}_{(-i)}^2), \boldsymbol{W}, \lambda^2\right] &\sim \mathrm{N}\left(\sum\nolimits_j (w_{ij}/w_{i+})\ \ln(\sigma_j^2),\ \lambda^2/w_{i+}\right), i = 1, \ldots, N,\\
\lambda^2 &\sim \mathrm{IG}\left(a,\ b\right),\\
c_i &= p_i/(1-p_i), i = 1, \ldots, N,\\
p_i &\overset{iid}{\sim} \mathrm{Beta}\left(\alpha,\ \beta\right),
\end{aligned} \tag{5}$$

with hyperprior parameters $a = b = 1, \alpha = \beta = 2$. $\mathrm{IG}(\cdot,\cdot)$ denotes the inverse gamma distribution and $\mathrm{Beta}(\cdot,\cdot)$ denotes the beta distribution. To simplify notation, we will denote $\tilde{\mu}_i = \sum_j (w_{ij}/w_{i+})\,\mu_j$. The p_i above requires explaining. Let $p_i = c_i/(1+c_i)$, then the full conditional for μ_i is

$$\left[\mu_i \mid y_i, \boldsymbol{\mu}_{(-i)}, \sigma_i^2, p_i\right] \sim \mathrm{N}\left(p_i y_i + (1-p_i)\tilde{\mu}_i,\ p_i\sigma_i^2\right). \tag{6}$$

Each p_i has an intuitive interpretation in our context: it is the parameter that controls the amount of smoothing at voxel i. If $p_i > 0.5$, more weight is placed on y_i, resulting in relatively less smoothing, while if $p_i < 0.5$ more weight is placed on $\tilde{\mu}_i$, resulting in relatively more smoothing. We will discuss our hyperprior choices in Section 4.

The BN and RH models

We now explain the differences between our proposed CWAS model and the two spatially adaptive models proposed by Brewer and Nolan (2007, BH) and Reich and Hodges (2008, RH), and our motivation for proposing a new model.

In the BN model, instead of using spatially varying error variances, the authors assume that $y_i = \mu_i + \epsilon_i$, with $\epsilon_i \sim \text{N}(0, \sigma^2)$. In addition, in the BN model the prior placed on μ_i is given by (1) with $b_{ij} = w_{ij}\{(\tau_i^2 + \tau_j^2)[\sum_{i\sim j}(\tau_i^2 + \tau_j^2)^{-1}]\}^{-1}$ and $\phi_i^2 = [\sum_{i\sim j}(\tau_i^2 + \tau_j^2)^{-1}]^{-1}$. Specifically, the conditional distribution for μ_i in (1) is given by

$$\left[\mu_i \mid \boldsymbol{\mu}_{(-i)}, \boldsymbol{\tau}^2\right] \sim \text{N}\left(\sum\nolimits_{i\sim j} \frac{\mu_j}{\tau_i^2 + \tau_j^2}\left(\sum\nolimits_{i\sim j} \frac{1}{\tau_i^2 + \tau_j^2}\right)^{-1}, \left(\sum\nolimits_{i\sim j} \frac{1}{\tau_i^2 + \tau_j^2}\right)^{-1}\right), \tag{7}$$

where $\boldsymbol{\tau}^2 = (\tau_1, \ldots, \tau_N)$.

In the BN model, the authors use an empirical Bayesian method to specify the hyperprior distributions on the τ_i^2s. They first fit the BYM model to their data to get raw estimates of the μ_is. Then, they set the mean of the prior for each τ_i^2 to be

$$\Lambda_i^{-1} = w_{i+}\left[\sum\nolimits_{j\in N(i)}(\mu_i - \mu_j)^2/w_{i+} - (\sum\nolimits_{j\in N(i)}(\mu_i - \mu_j)/w_{i+})^2\right]/2\,,$$

and implement their spatially adaptive smoothing model to get the final estimate of μ_i. When we apply the BN model in our simulation study and to our data, we set $\tau_i^2 \sim \text{IG}(2, \Lambda_i^{-1})$, for each i.

Reich and Hodges (2008) take a different approach by setting b_{ij} and ϕ_i^2 in (1) to $w_{ij}\tau_j^{-1}/\sum_{i\sim k}\tau_k^{-1}$ and $\tau_i/\sum_{i\sim j}\tau_j^{-1}$, respectively. In the RH model the conditional distributions of μ_i, $i = 1, \ldots, N$ are

$$\left[\mu_i \mid \boldsymbol{\mu}_{(-i)}, \boldsymbol{\tau}^2\right] \sim \text{N}\left(\sum\nolimits_{i\sim j}\left(\tau_j^{-1}/\sum\nolimits_{i\sim k}\tau_k^{-1}\right)\mu_j,\ \tau_i/\sum\nolimits_{i\sim j}\tau_j^{-1}\right). \tag{8}$$

Similar to our CWAS model, in the RH model the authors assume that $y_i = \mu_i + \epsilon_i$, with $\epsilon_i \sim \text{N}(0, \sigma_i^2)$. Both the σ_i^2s and τ_i^2s vary spatially with CAR priors placed on their logarithms:

$$\begin{aligned}\left[\ln(\sigma_i^2) \mid \ln(\boldsymbol{\sigma}^2_{(-i)}), \boldsymbol{W}, \lambda^2\right] &\sim \text{N}\left(\sum\nolimits_j (w_{ij}/w_{i+})\ln(\sigma_j^2),\ \lambda^2/w_{i+}\right),\\ \left[\ln(\tau_i^2) \mid \ln(\boldsymbol{\tau}^2_{(-i)}), \boldsymbol{W}, \gamma^2\right] &\sim \text{N}\left(\sum\nolimits_j (w_{ij}/w_{i+})\ln(\sigma_j^2),\ \gamma^2/w_{i+}\right),\end{aligned} \tag{9}$$

where λ^2 and γ^2 are both assumed to follow an inverse Gamma distribution, IG(1,1).

Our model is different from both the BN model and the RH model in the definitions of the b_{ij} and ϕ_i^2 in (1). We believe that compared to the BN and RH models our model offers a more intuitive interpretation. The model parameter p_i for voxel i is the weight placed on the data y_i and controls the amount of smoothing in the CWAS model at voxel i.

2.3 Loss function

The goal of using fMRI for pre-surgical planning is to identify which voxels did not activate or deactivate and are safe to resect. After fitting our CWAS model to data, we

use a Bayesian decision theoretic approach to classify voxels into two classes, null and non-null (those voxels that are either activated or deactivated). A typical fMRI data analysis places emphasis on only controlling the type I error. However, in pre-surgical fMRI planning, surgeons are more concerned with false negatives since damage to false negative regions of the brain may lead to irreversible deficits in function. Therefore, we propose a novel loss function that explicitly controls both false negatives and false positives, while at the same time allowing false negatives and false positives to be treated asymmetrically in the loss function. Our loss function is a generalization of a loss function proposed by Müller et al. (2007). Let μ_i indicate the mean intensity for voxel i, $r_i \in \{0,1\}$ denote the true binary state of voxel i (0 for null, 1 for non-null), and $\delta_i \in \{0,1\}$ represent the estimated state. Let $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_N)$ and $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_N)$ be the vectors corresponding to all voxels and let k_1, k_2 and t be positive weights. Then our proposed loss function is defined as follows:

$$\mathrm{L}(\boldsymbol{\mu}, \boldsymbol{\delta}) = \sum\nolimits_i \left[-r_i\delta_i - (1-r_i)(1-\delta_i) + k_1 r_i(1-\delta_i) + k_2(1-r_i)\delta_i + t\delta_i\right].$$

The terms in this loss function indicate: (i) two types of gains corresponding to correct classification of an activated or deactivated voxel; $-r_i\delta_i$, and of a null voxel; $-(1-r_i)(1-\delta_i)$; (ii) two types of losses corresponding to incorrect classification of an activated or deactivated voxel; $k_1 r_i(1-\delta_i)$, and of a null voxel; $k_2(1-r_i)\delta_i$; and (iii) a penalty term for the total number of discoveries ensuring parsimony, $\sum_i t\delta_i$. Above, k_1 is a weight for incorrectly classified non-null voxels (false negatives) while k_2 is a weight for incorrectly classified null voxels (false positives). When $k_1 > k_2$ more weight is given to incorrectly classified activated voxels relative to null voxels.

Given the data, the optimal decision rule is obtained when the posterior expected loss

$$\begin{aligned}\mathbb{E}\left(\mathrm{L}(\boldsymbol{\mu}, \boldsymbol{\delta} \mid \boldsymbol{Y})\right) &= \sum\nolimits_i \left[-(2+k_1+k_2)\delta_i \Pr(r_i = 1 \mid \boldsymbol{Y}) + (1+k_1)\Pr(r_i = 1 \mid \boldsymbol{Y})\right. \\ &\quad \left. + (1+k_2+t)\delta_i - 1\right]\end{aligned}$$

is minimized. The optimal decision is

$$\delta_i^{\dagger} = \mathrm{I}\left[\Pr(r_i = 1 \mid \boldsymbol{Y}) \geq (1+k_2+t)/(2+k_1+k_2)\right].$$

However, we do not explicitly define $\Pr(r_i = 1 \mid \boldsymbol{Y})$ in our model. Instead, we consider for each voxel i

$$f(m_i) = f\left(\left|\mathbb{E}(\mu_i \mid \boldsymbol{Y})/\sqrt{\mathrm{Var}(\mu_i \mid \boldsymbol{Y})}\right|\right)$$

where f is some monotone function and m_i can be consider as the strength of a voxel being non-null (i.e., either activated or deactivated); with $m_i = 0$ if $r_i = 0$ and $m_i > 0$ when $r_i = 1$ (see Müller et al. (2007) for details). Our loss function is now

$$\begin{aligned}\mathrm{L}(\mathbf{m}, \boldsymbol{\delta}) &= \sum\nolimits_i \{-f(m_i)\delta_i - [1-f(m_i)](1-\delta_i) + k_1 f(m_i)(1-\delta_i) \\ &\quad + k_2[1-f(m_i)]\delta_i + t\delta_i\}\end{aligned}$$

where $\mathbf{m} = (m_1, m_2, \ldots, m_N)$ and the optimal decision is

$$\delta_i^{\star} = \mathrm{I}\left[f(\widehat{m}_i) \geq (1+k_2+t)/(2+k_1+k_2)\right]$$

and $\widehat{m}_i$ is the estimated posterior of m_i. For the function f, a reasonable choice is that by Gross and Binder (2014), that is, $f(m_i) = m_i/q_\alpha(\mathbf{m})$ where q_α is the $(1-\alpha)$-quantile function of its argument. By choosing an appropriate level of α, one can guard against outliers. In our application, there does not appear to be any outliers, and so we choose $\alpha = 0$. That is, $q_0(\mathbf{m}) \equiv \max(\mathbf{m})$. Another appropriate choice would be $\alpha = 0.01$. The decision rule of our proposed loss function only depends on the values of the constants k_1, k_2 and t. Once fixed, we have a uniform decision rule that we can apply across different data sets.

3 Posterior estimation

3.1 The CWAS model

We now discuss posterior estimation for the parameters in our model. We implement a hybrid Metropolis-within-Gibbs algorithm to sample the parameters from their full conditionals. The full conditionals for $\mu_i,\ i = 1, \ldots, N$ and λ^2 are available in closed form and are respectively the normal distribution given in (6) and an inverse gamma distribution

$$\left[\lambda^2 \mid \boldsymbol{\sigma}^2\right] \sim \text{IG}\left(a + 0.5(N-1),\, b + 0.5\sum\nolimits_{i\sim j}\left[\ln(\sigma_i^2) - \ln(\sigma_j^2)\right]^2\right).$$

The prior and posterior for both σ_i^2 and c_i are not conjugate pairs. The full conditionals, up to a constant of proportionality, are:

$$\pi(\sigma_i^2 \mid y_i, \boldsymbol{\sigma}^2_{(-i)}, \boldsymbol{\mu}, \lambda^2) \propto \pi(y_i \mid \mu_i, \sigma_i^2)\pi(\sigma_i^2 \mid \boldsymbol{\sigma}^2_{(-i)}, \lambda^2)\pi(\boldsymbol{\mu} \mid \boldsymbol{c}, \boldsymbol{\sigma}^2), \tag{10}$$

$$\pi(c_i \mid \boldsymbol{c}_{(-i)}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \propto \pi(c_i)\pi(\boldsymbol{\mu} \mid \boldsymbol{c}, \boldsymbol{\sigma}^2), \tag{11}$$

where the prior distribution of c_i is induced by the prior distribution on p_i. The prior density of c_i is

$$\pi(c_i) = c_i^{\alpha-1}(1+c_i)^{-(\alpha+\beta)}/\text{B}(\alpha,\beta),$$

where $\text{B}(\cdot,\cdot)$ denotes the Beta function. We remark that although the c_is are a priori independent, they are spatially correlated in the posterior. The full conditional of c_i is

$$\pi(c_i \mid \boldsymbol{c}_{(-i)}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) \propto c_i^{-\frac{1}{2}} \exp\left\{-\left(\mu_i - \tilde{\mu}_i\right)^2/(2c_i\sigma_i^2)\right\} c_i^{\alpha-1}(1+c_i)^{-(\alpha+\beta)}.$$

Note that this full conditional depends on $\tilde{\mu}_i = \sum_j (w_{ij}/w_{i+})\mu_j$—thus borrowing information from neighboring voxels. Furthermore, the posterior of c_i depends on μ_i and σ_i^2 which are explicitly dependent on their neighbors (both a priori and a posteriori). Thus, the c_is (and p_is) are a posteriori correlated and borrow strength implicitly from their neighboring values. In the Discussion section of the paper, we illustrate how one can modify the priors on the p_is so that they are spatially dependent a priori, which may be desirable.

The full conditionals in (10) and (11) are intractable due to the fact that $\pi(\boldsymbol{\mu} \mid \boldsymbol{c}, \boldsymbol{\sigma}^2)$ is intractable. By Brook's Lemma (Brook, 1964), the prior of $\boldsymbol{\mu}$ is guaranteed

to exist. However, with our specification of the conditional priors on the μ_is in (5), the prior of $\boldsymbol{\mu}$ does not have a tractable density. To overcome this issue, we use the pseudo-likelihood approach (Besag, 1975) to approximate the prior of $\boldsymbol{\mu}$, $\pi(\boldsymbol{\mu} \mid \boldsymbol{c}, \boldsymbol{\sigma}^2)$. The pseudo-likelihood approximation was originally developed for spatial models by Besag (1975) motivated by the computational intractability of the true likelihood. Besag showed that parameter estimation using the pseudo-likelihood approximation is efficient for simple Gaussian fields (Besag, 1977). The pseudo-likelihood approximation proposed by Besag (1975) for $\boldsymbol{\mu}$ is formulated as the product of all the full conditionals $\pi(\mu_i \mid \cdot)$:

$$\pi(\boldsymbol{\mu} \mid \boldsymbol{c}, \boldsymbol{\sigma}^2) \approx \prod_{i=1}^{N} (2\pi c_i \sigma_i^2)^{-\frac{1}{2}} \exp\left\{ -\left(\mu_i - \tilde{\mu_i}\right)^2 / (2c_i\sigma_i^2) \right\},$$

where $\boldsymbol{c} = (c_1, \ldots, c_N)^T$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_N^2)^T$. Since the full conditionals of σ_i^2 and c_i, $i = 1, \ldots, N$, do not have closed forms, we draw samples from their full conditionals using the Metropolis–Hastings algorithm (Hastings, 1970). Note that only the estimation of the c_is and the σ_i^2s rely on the pseudo-likelihood approximation.

3.2 The BN and RH models

Brewer and Nolan (2007) also used the pseudo-likelihood approximation to the prior of $\boldsymbol{\mu}$ in their model. In their case, the prior for $\boldsymbol{\mu}$ is approximated as

$$\pi(\boldsymbol{\mu} \mid \boldsymbol{\tau^2}) \approx \prod_{i=1}^{N} \left(\frac{2\pi}{\sum_{i\sim j} \left(\tau_i^2 + \tau_j^2\right)^{-1}} \right)^{-\frac{1}{2}} \exp\left\{ - \frac{\left(\mu_i - \sum_{i\sim j} \frac{\mu_j}{\tau_i^2+\tau_j^2} \Big/ \sum_{i\sim j} \frac{1}{\tau_i^2+\tau_j^2} \right)^2}{2\left[\sum_{i\sim j} \left(\tau_i^2 + \tau_j^2\right)^{-1}\right]^{-1}} \right\}.$$

Reich and Hodges (2008) use the following joint prior distribution for $\boldsymbol{\mu}$:

$$\pi(\boldsymbol{\mu} \mid \boldsymbol{\tau}^2) \propto |Q|^{\frac{1}{2}} \exp\left\{ -0.5 \sum\nolimits_{i\sim j} \left(\mu_i - \mu_j\right)^2 / (\tau_i \tau_j) \right\},$$

where the matrix Q has entries $Q_{ii} = \sum_{i\sim j} \tau_{ij}^{-2}$ and $Q_{ij} = -I(i \sim j)\tau_{ij}^{-2}$, while the determinant of Q is calculated as the product of all positive eigenvalues of the non-negative definite matrix Q. For our Z-statistic image with approximately 47000 voxels (after skull stripping and exclusion of non-brain voxels), computing the determinant of Q is computationally infeasible; therefore, when we apply the RH model to our Z-statistic image, we implement a pseudo-likelihood approach to approximate the prior of $\boldsymbol{\mu}$ with

$$\pi(\boldsymbol{\mu} \mid \boldsymbol{\tau^2}) \approx \prod_{i=1}^{N} \left(\frac{2\pi\tau_i}{\sum_{i\sim j} \tau_j^{-1}} \right)^{-\frac{1}{2}} \exp\left\{ - \left(\mu_i - \sum\nolimits_{i\sim j} \frac{\tau_j^{-1}\mu_j}{\sum_{i\sim k} \tau_k^{-1}} \right)^2 \Big/ \left(\frac{2\tau_i}{\sum_{i\sim j} \tau_j^{-1}} \right) \right\}.$$

4 Simulation studies

We conducted simulation studies to investigate the performance of the proposed CWAS model. In the simulation studies, we (i) investigate sensitivity to different prior specifications on the p_is; (ii) demonstrate the relationship between the number of correctly

classified activated voxels and the thresholds used to detect activation regions for the simulated fMRI data with different signal to noise ratios; and (iii) compare the performance of our proposed CWAS model with that of the BN and RH models.

Center Voxel	[20,40,10]	[36,50,18]	[31,35,20]	[53,29,25]	[40,40,30]	[46,25,33]
Effect Size	10	8	6	4	2	1
Radius	2	4	1	2	4	1

Table 1: Locations, effect sizes and radii of the 6 activated regions in the simulation studies.

For simulation purposes, we used the neuRosim package described in Welvaert et al. (2011) to simulate a set of four-dimensional fMRI data. The detailed settings used for data generation are elaborated as follows. A boxcar block design was used for the stimulus function, with each on/off episode lasting 20 seconds. Repetition time (TR) is set to two seconds with a total of 100 scans in each simulated data set. Thus each simulated experiment lasts 200 seconds, with activation onset at 21, 61, 101, 141, and 181 seconds. The stimulus function is convolved with a gamma hemodynamic response function (HRF). The brain mask for the simulated data is chosen to be the same as that of the patient described in Section 5. The brain mask is a three-dimensional $64 \times 64 \times 40$ binary array indicating whether an element in the array corresponds to a voxel in the brain. There are six spherical activation regions with different sizes and HRF signal magnitudes. Details of the six activation regions are shown in Table 1. The baseline signal is set to 250 and the signal fading rate is set to 0.01 as suggested by Welvaert et al. (2011). The noise is a mixture of Gaussian white noise (30%), and spatial noise (70%) with default auto-correlation coefficient equal to 0.75. We simulated 50 data sets for each signal-to-noise ratio (SNR) considered (SNR = 1, 2 and 3). The average SNR is defined as SNR$= \bar{S}/\sigma_N$, where $\bar{S}$ represents the average magnitude of the signal, and σ_N stands for the standard deviation of the noise (Krüger and Glover, 2001). Z-statistic images were created using FEAT within the FSL (Smith et al., 2004) software package with spatial smoothing turned off in the preprocessing stage. Based on these settings, each Z-statistic image contains a total of 46932 voxels, 328 of which are truly active.

The 150 simulated data sets were then fitted with the CWAS, RH, and BN models. For all three models, we ran the MCMC algorithm for 150,000 iterations with the first 100,000 iterations discarded as burn-in. The RH model encountered convergence problems that might be attributable to (i) the large data size, or (ii) the pseudo-likelihood approximation. However, when the RH model did converge, our model always converged faster. We visually assessed the traceplots based on which we concluded our model converged faster. Convergence was further verified using the Gelman–Rubin convergence diagnostic statistic (Gelman and Rubin, 1992; Brooks and Gelman, 1998). For cases when the RH model did not converge, we reran the MCMC simulations using a different starting value so that all models were compared using the same 150 simulated data sets.

To investigate sensitivity to different prior specifications under the CWAS model, we considered three Beta distributions, Beta(2,2), Beta(1,3) and Beta(3,1), as the hyperprior distribution of the p_is. We also attempted to control the amount of smoothing

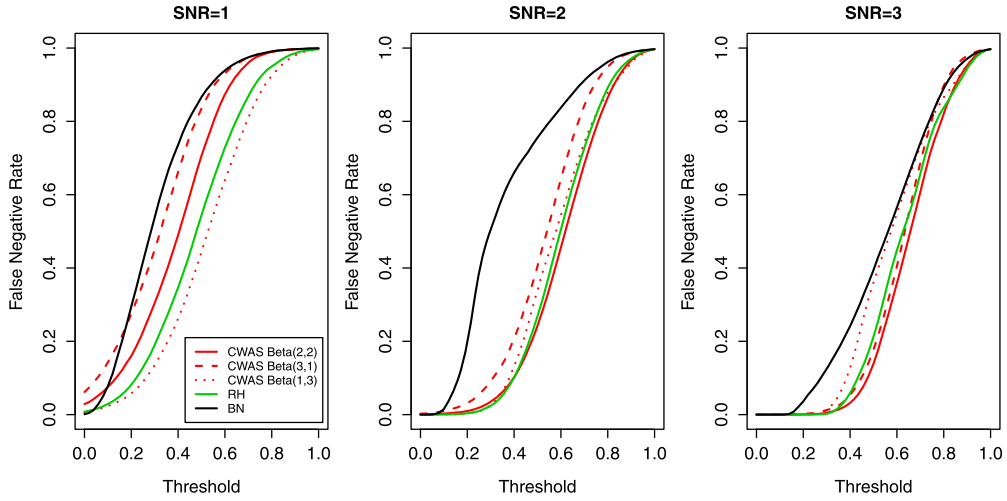

Figure 1: False negative rate vs threshold in the simulation studies. For each signal to noise ratio, we take the average of the false negative rate over 50 simulations using different Beta prior distributions for the p_is in our CWAS model. For comparison we also include results using the BN and RH models. Threshold varies from 0 to 1.

in the RH model by implementing different hyperpriors on λ^2 or γ^2 (see (9)). However, smoothing was rather insensitive to the prior information unless we specified a very tight prior. A similar phenomenon was observed for the BN model. We applied our proposed loss function and summarized the false negative rate (FNR) for all three models. Instead of using a fixed threshold derived by minimizing the posterior expected loss and holding the constants k_1, k_2 and t in the loss function fixed, we varied the threshold from 0 to 1 to demonstrate the trade-off between threshold and true positive counts. For each SNR, Figure 1 shows the average FNR over the 50 simulated data sets versus threshold. Each curve denotes a different amount of smoothing for the CWAS model. Beta(2,2), Beta(3,1) and Beta (1,3) are different priors for the smoothing parameters p_i, which represents the weight assigned to the data point y_i: a larger p_i implies less smoothing at voxel i. As shown in Figure 1, the curve corresponding to Beta(1,3) results in the largest amount of smoothing and has the lowest FNR compared to the other two priors at a SNR of 1. On the other hand, the curve corresponding to Beta(1,3) has the largest FNR when the SNR is 3. This is consistent with our intuition that less smoothing is necessary as the SNR increases. We also note that our model has a consistently lower FNR than the other two models for a large range of thresholds.

We compared the smoothed Z-statistic images derived from the three models applied on one selected simulated data set. Figure 2 shows the comparison between the posterior estimates of the mean intensities from the three spatially adaptive smoothing models when SNR = 2. Similar results, not shown, were obtained for signal-to-noise ratios equal to 1 and 3 as well. The top row of Figure 2 displays the true activation regions in four sagittal slices of the brain, while the second row presents the simulated data with noise

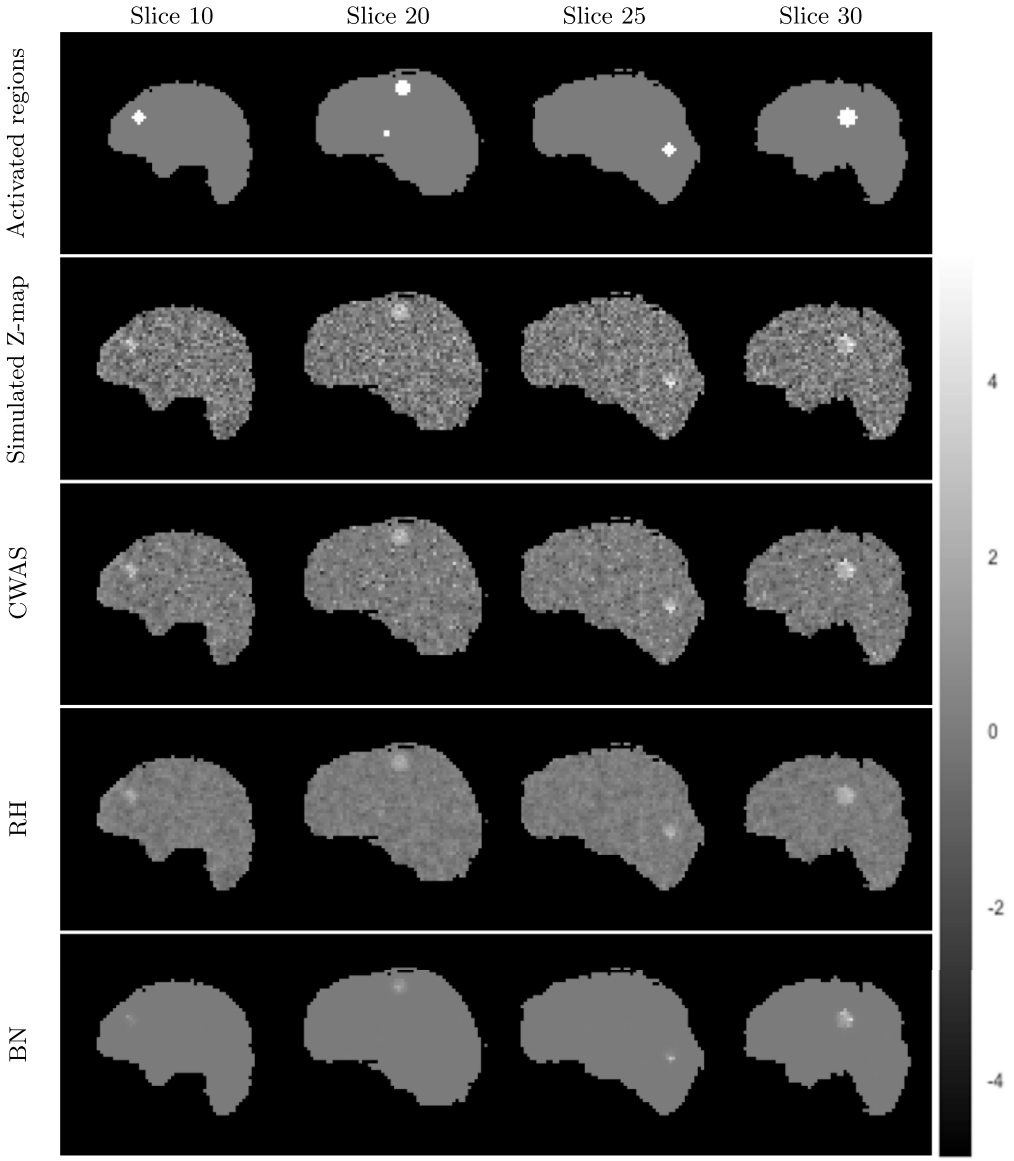

Figure 2: Comparison of the posterior estimates of the mean intensities from the three different spatially adaptive smoothing models when SNR = 2. The top row shows the truly activated regions in four sagittal slices of the brain (one of the six simulated regions does not appear in these four slices). The second row shows the simulated Z-statistic images (with no smoothing). The third through fifth rows show the marginal posterior means of $\boldsymbol{\mu}$ from the three spatially adaptive smoothing models.

added. By comparing row 3 through 5, it is evident that the amount of smoothing varies among the three models. The BN model severely smoothes the data. The RH model also tends to oversmooth the data. Our model smoothes the simulated data the least and it achieves the goal of balancing smoothing and retaining relatively sharp boundaries between activated and non-activated areas.

5 Application to pre-surgical fMRI data

The motivation of this work is to establish and test the performance of a novel, spatially adaptive smoothing model for pre-surgical fMRI data. In this section, we illustrate the application of our proposed CWAS model to the fMRI data from two brain tumor patients and compare it with the RH and BN models. Both patients were pre-surgically mapped by fMRI to determine the peri- and potentially intra-tumoral regions of the brain vital for speech and language functions in order to optimize (a) access to the tumor, (b) intra-operative ESM at a limited number of predefined cortical sites and (c) extent of resection without inflicting new aphasic deficits.

The first patient's tumor was located in proximity to the so-called dorsal stream of speech and language while the second patient's tumor was in proximity to the so-called ventral stream (Hickok and Poeppel, 2007). Consequently, two different fMRI paradigms were applied, specifically challenging functions of the dorsal and ventral stream (see below). A 30 seconds ON/OFF-boxcar block design was used in both cases, and task speed was adjusted to the individual's optimal performance level. Data on the first patient are from a 32 year old, right-handed woman who had initially experienced a seizure and was then brought into the emergency room, still with difficulties speaking. In particular, she had persistent problems repeating phonemically challenging phrases. In the fMRI paradigm, the patient was therefore instructed and cued to alternate between the silent recitation of challenging tongue twister phrases (ON) and covert repetition of the unchallenging, rhythmic phoneme sequence "tock-tock-tock" (OFF). After partial resection, the tumor, in the left insular and inferior frontal lobe, was classified as an oligodendroglioma with anaplastic components (see the structural MRI scan in the left panel of Figure 3 and Figure 4). Data on the second patient are from a 62 year old, right-handed woman who presented difficulties with word finding, comprehension and reading but who did not display signs of agraphia. In the fMRI paradigm, the patient was therefore asked to silently read non-final embedded clause sentences (ON) alternating with consonant strings (OFF). After partial resection, the tumor, located in the left middle and inferior temporal gyrus, was classified as a glioblastoma multiforme (GBM) with intra-tumoral hemorrhages (see Figure 5).

Both fMRI studies were recorded on a 3 Tesla TimTrio scanner (TQ engine, 32 channel head coil; Siemens Medical Solutions, Erlangen) using gradient-echo (GE) echo-planar imaging (EPI) at the following parameters: TR = 3000 ms, TE = 30 ms, echo spacing = 0.69 ms, GRAPPA acceleration factor 2, $n = 160$ volumes in time, $64 \times 64 \times 40$ voxels in space ($3.00 \times 3.00 \times 3.00$ mm^3 resolution, with an additional 15% inter-slice gap). Geometric distortions were corrected using an additional GE fieldmap scan. High-resolution anatomical T2-weighted FLAIR- (fluid-attenuated inversion recovery) and

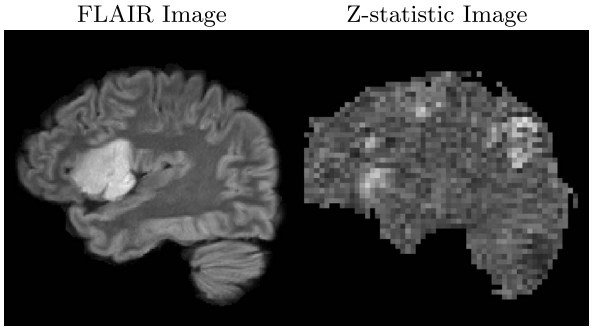

Figure 3: Pre-surgical fMRI imaging. The left panel shows a FLAIR (fluid-attenuated inversion recovery) image where the tumor of the first patient is evident. The right panel shows the unsmoothed Z-statistic image. The Z-statistic image generated by mass-univariate general linear modeling is related to the probability of brain activation.

T1-weighted MPRAGE (magnetization-prepared rapid gradient echo) sequences were recorded as well for intra-operative neuro-navigation, cortical surface reconstruction and the display of fMRI results (cf. Figures 4 and 5). A distortion-corrected 3-dimensional Z-statistic image was estimated in native GE-EPI space ($64 \times 64 \times 40$ voxels) for each patient using the FEAT analysis tool v.6.00 (with convolution of the design matrix by the canonical double-gamma hemodynamic response function, HRF, and inclusion of temporal derivatives) in FSL 5.0.7 (`http://www.fmrib.ox.ac.uk/fsl/`, Smith et al. (2004); Woolrich et al. (2009)). Spatial smoothing was turned off during the preprocessing stage, motion correction was performed and data were temporally filtered using a high-pass cutoff of 90 seconds (default Gaussian-weighted least-squares straight line fitting). A single sagittal slice of the resulting Z-statistic image of the first patient is shown in the right panel of Figure 3.

In the following analysis, we fit our proposed CWAS model to these Z-statistic images. Bayesian decision theory, using the loss function defined in Section 2.3, is subsequently used to identify activated and deactivated voxels. For comparison, results from the RH and BN models are shown as well (cf. Figures 4 and 5). We ran our model for 50,000 iterations after a burn-in of 100,000 (total iterations 150,000). The MCMC algorithm takes approximately 1 hour CPU time on an iMAC with a 3.2 GHz Intel Core i5 processor with 16 GB of memory.

Based on our simulation studies, we use a Beta(2,2) prior for the parameters $p_i, i = 1, \ldots, N$, since this prior showed consistently good performance in identifying activated regions over a large range of thresholds for SNR = 2, 3. Welvaert et al. (2011) demonstrated that the estimated SNR from real fMRI data is around 3.87. Next, we want to determine a threshold to identify activated and deactivated voxels. Since, in pre-surgical fMRI, misclassification of activated or deactivated voxels is more important than false positives, we set $k_1 = 11,\ k_2 = 1,\ t = 1$ in the posterior expected loss function. This results in a loss that penalizes false negatives 11 times more heavily than false positives and was determined by expert opinion (A.J. Bartsch) and is concordant with the

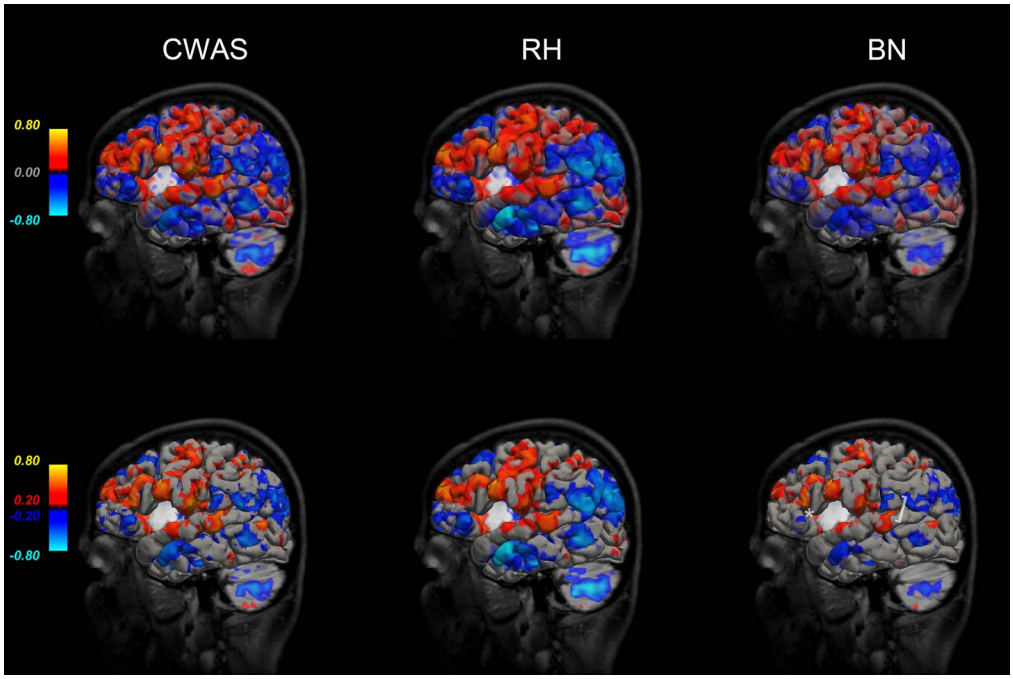

Figure 4: Results for patient 1: CWAS, RH and BN models. The top row shows results on the loss function scale projected onto the cortical surface. In the bottom row, the loss function has been truncated below by 0.2 and above by –0.2—indicating those regions that survived thresholding. Values above 0.8 are mapped to 0.8 and values below –0.8 are mapped to –0.8 to give a better dynamic range of colors. Red to yellow denote increasingly strong activation. Dark blue to bright blue denote increasingly strong deactivation. In the lower right panel, the tip of the pars triangularis of the inferior frontal gyrus is marked with a white asterisk * and the peri-Sylvian supramarginal gyrus (SMG or area Spt) is marked with a slightly rotated white bracket].

results obtained by ESM (see below). We classify any voxel whose posterior expected mean intensity is positive, i.e., $\mathbb{E}(\mu_i \mid \boldsymbol{Y}) > 0$, and for which $f(m_i) = m_i/q_0(\mathbf{m}) \geq 0.2$ as activated. Conversely, those voxels for which the posterior expected mean intensity is negative, i.e., $\mathbb{E}(\mu_i \mid \boldsymbol{Y}) < 0$, and $f(m_i) \geq 0.2$ are considered deactivated.

Figures 4 and 5 display the activated and deactivated regions, for the two patients, projected onto the cortical surface identified by the CWAS, RH and BN models. Reconstruction of the cortical surface, surface editing to visualize the tumor, transformation/projection of the statistical maps onto the pial surface and the underlying volume space was performed using FreeSurfer v.5.3.0 (`http://surfer.nmr.mgh.harvard.edu/`; Fischl (2012)) and its recon-all, reg-feat2anat and feat2surf scripts. Left hemispheres are visualized using Freeview (part of FreeSurfer). A minimal amount of smoothing was applied to the pial surface and surface projection of the data, respectively, and is the re-

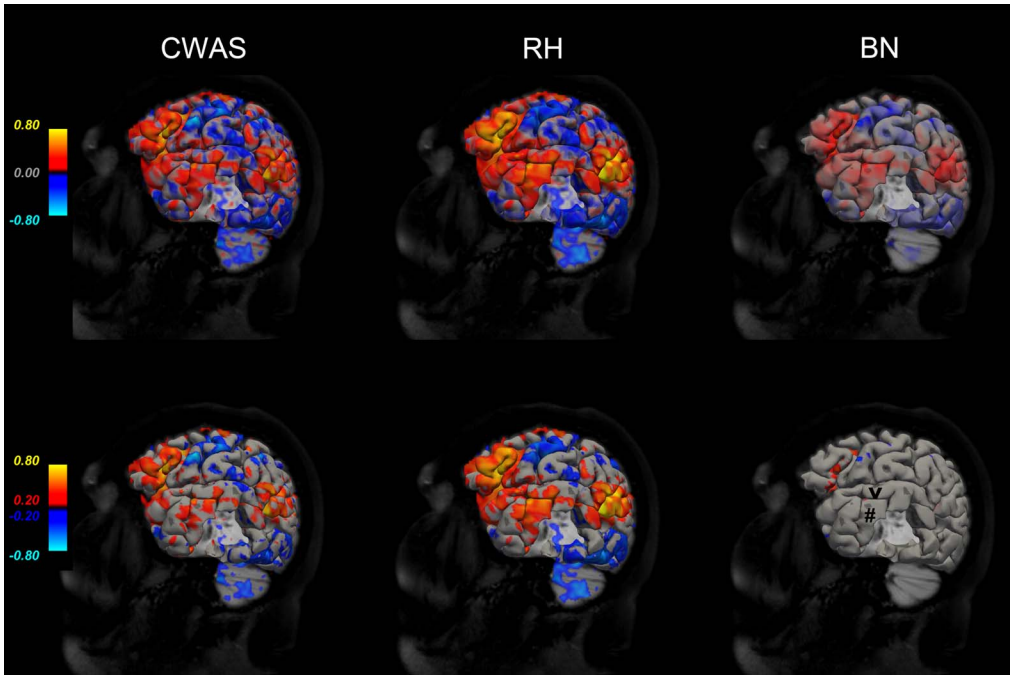

Figure 5: Results of the CWAS, RH and BN models for patient 2 (cf. Figure 4 for further explanations). In the lower right panel, the superior temporal sulcus (STS or t1) is marked with a black arrowhead and the middle temporal gyrus (MTG or T2) is marked with a black #.

sult of a spline interpolation of the data into the high-resolution ($1.00 \times 1.00 \times 1.00$ mm^3) anatomical volume space. The tumors are made visible through an edited "hole", i.e., a topological defect, of the cortical surface.

Although surgeons are more concerned with activated regions, in Figures 4 and 5 we also present deactivation regions because in brain tumor patients even decreases in the BOLD signal may be related to brain activation (Fujiwara et al., 2004; Hsu et al., 2004; Ulmer et al., 2004; Bartsch et al., 2006). As a result, deactivated voxels should be considered in the loss function (although we acknowledge that "paradoxical" BOLD signal decreases may often represent a rather slightly, i.e., by the order of a few seconds, increased delay of the peri- or intra-tumoral HRF and are then less likely to reflect true brain activation in fMRI designs of longer blocks such as 30 seconds modeled with temporal derivatives).

Activated and deactivated areas detected by the BN model are clearly less extensive than those obtained by either the CWAS or RH model. While the BN model still performed reasonably in the case of the first patient (Figure 4), it clearly did not in the case of the second patient (Figure 5). On the other hand, activated and deactivated regions detected by the RH model are obviously larger and "smoother", i.e., more con-

tinuously distributed over surface vertices/volume voxels, than those obtained under the CWAS model. Under the RH model, the spatial extent exceeded beyond what is expected according to expert opinion (A.J. Bartsch) in both clinical cases.

Furthermore, these impressions and superiority of the CWAS model were confirmed by intra-operative ESM in both patients—in the first patient, there was fading evidence for activation according to the RH model at the tip of the pars trigangularis of the inferior frontal gyrus (IFG or F3) towards the Sylvian fissure which was eloquent based on electrical stimulation (i.e., ESM evoked speech arrest at this location) and active according to the CWAS results. Compared to the BN model, CWAS results seemed superior at the peri-Sylvian supramarginal gyrus (SMG) of the inferior parietal lobule (P2) which was not stimulated by ESM in this case but is a known speech center of the dorsal stream, i.e., the Sylvian fissure at the parietotemporal boundary (or Spt) area (cf. Hickok and Poeppel (2007)).

In the second patient, the BN model essentially failed to reveal peri- or intra-tumoral fMRI activations. Here, the RH model suggested high probabilities of activation in the middle temporal gyrus (MTG or T2) which were only confirmed for the area around the superior temporal sulcus (STS or t1) by ESM at approximately the line indicated by the CWAS statistical image thresholded at ≥ 0.20 (cf. lower left panel in Figure 5).

Evidence for intra-tumoral fMRI activations was low according to all three models in the first patient and essentially absent on the thresholded CWAS, RN and BH maps (cf. Figure 4, lower panel). This was confirmed by ESM which did not lead to speech arrest or other speech and language disturbances during awake craniotomy upon stimulation of the tumor itself. In this case, the resection was limited by proximity to and infiltration of the arcuate fasciculus by the tumor (not shown; cf. Bartsch et al. (2014), with Figure 23.14 there for illustration and further discussion of eloquent fiber pathways that can be tracked by diffusion-weighted imaging and are similarly important for pre-surgical planning). On the contrary, there was evidence for intra-tumoral activations in the second patient according to the CWAS and RH model which survived the thresholding only in CWAS (cf. Figure 5). ESM of the tumor in this case did, in fact, worsen naming, comprehension and reading performance in the patient and eventually lead to speech arrest which limited the surgical resection. Function in tumor is a known phenomenon to occur primarily in highly malignant tumors (like GBM in the second patient; Ojemann et al. (1996)). Note that the correspondence between ESM and the CWAS, RH and BN maps was assessed after the surgery and was not used for the actual medical decision making.

We conclude this section with a discussion on the quality of the model parameter estimates. Given the large number of parameters in the CWAS model, one may wonder whether there is enough information in the data to inform on all of the parameters. The answer is yes. Patient 1's dataset consists of approximately 47000 voxels and patient 2's dataset consists of approximately 54000 voxels. For each voxel, there are three parameters to estimate, μ_i, p_i, and σ_i^2. Although there are many more parameters to estimate than voxels, there is information in the data to estimate all of the parameters. First, the Gelman–Rubin diagnostic statistic shows that the MCMC chains for the μ_is, σ_i^2s, and p_is all have converged. For each parameter, the Gelman–Rubin R statistic (Gelman

	μ_is	p_is	σ_i^2s
Patient 1	(0.999952, 1.000414)	(0.999954, 1.027688)	(0.999953, 1.000770)
Patient 2	(0.999953, 1.000486)	(0.999950, 1.000052)	(0.999950, 1.000466)

Table 2: The ranges of the Gelman–Rubin diagnostic statistic for the μ_is, p_is, and σ_i^2s.

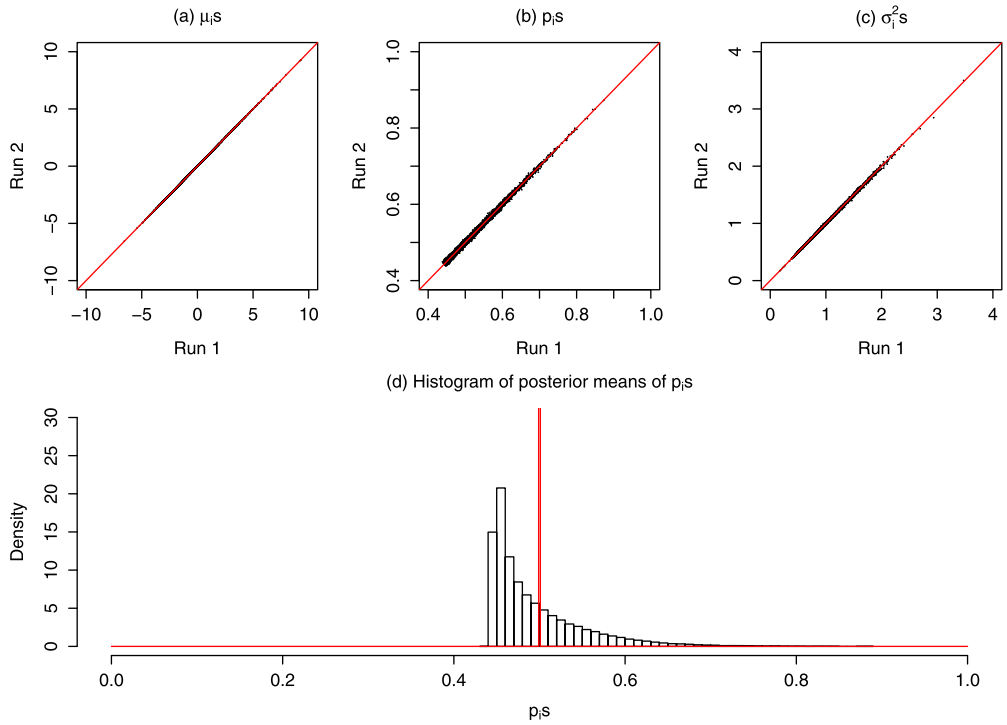

Figure 6: (a)–(c) Scatterplots of posterior mean estimates from two independent MCMC runs from patient 1; (d) Histogram of the posterior estimates of p_is with red line indicating the sampling distribution of the posterior means of p_is from patient 1 run 1.

and Rubin, 1992; Brooks and Gelman, 1998) is calculated from five MCMC simulations with different starting values. The ranges of the Gelman–Rubin diagnostic statistic for the μ_is, p_is, and σ_i^2s are shown in Table 2. All ranges are below the suggested threshold of 1.2. Note that we used the univariate Gelman–Rubin diagnostic statistic. Although there is a multivariate version, it is not possible to use it in our application as we have far too many parameters.

Figure 6 (a)–(c) and Figure 7 (a)–(c) display scatterplots of these parameters from two independent MCMC runs from patient 1 and patient 2, respectively. These figures further corroborate the conclusions from the Gelman–Rubin statistic. In Figure 6 (d) and Figure 7 (d), we also show a histogram of all of the posterior means of the p_is for each patient. The red curves, overlaying the histograms, are the theoretical distribution

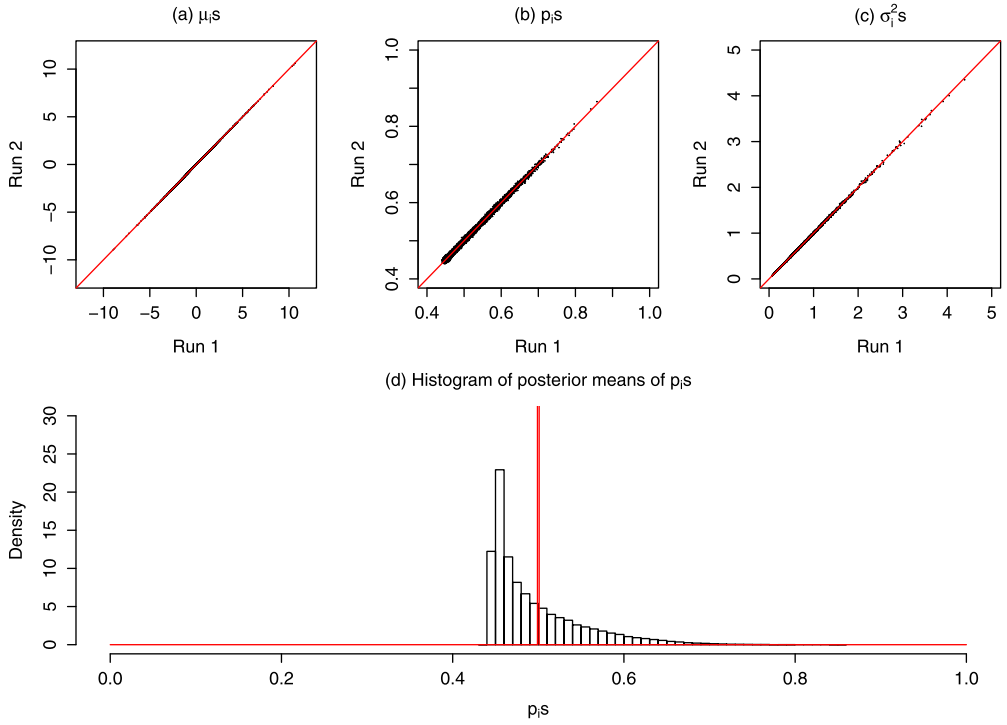

Figure 7: (a)–(c) Scatterplots of posterior mean estimates from two independent MCMC runs from patient 2; (d) Histogram of the posterior estimates of p_is with red line indicating the sampling distribution of the posterior means of p_is from patient 2 run 1.

of the posterior expectation of the p_is under the assumption that the data are completely uninformative and the p_is are independent. It is evident from these two figures that the data are informing the p_is, even though we assume that they are independent a priori. Figure 8 shows a histogram of the MCMC standard errors of the μ_is and a scatterplot of MCMC standard errors versus the marginal posterior mean estimates of the μ_is for each patient. MCMC standard errors were calculated using the "mcmcse" package in R (Flegal and Hughes, 2012) with the default batch means method (Flegal et al., 2008).

6 Discussion

In this paper, we propose an alternative model to the NP-Potts model developed by Johnson et al. (2013) for analyzing pre-surgical fMRI data. We have developed a novel spatially adaptive smoothing model (CWAS) that allows the data to determine where and how much smoothing should occur in the Z-statistic image. We have incorporated a Bayesian decision theoretical approach to classify the voxels into activation, deactivation and null states. Simulating from the posterior of the CWAS model is much more com-

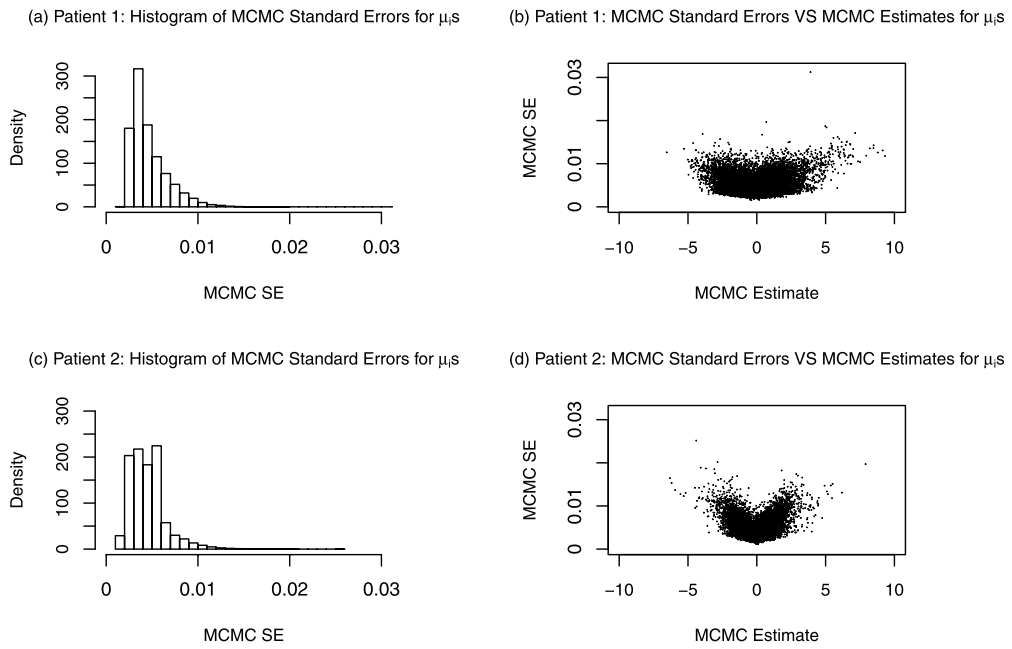

Figure 8: MCMC standard errors (SE) for the μ_is: (Left column) Histograms of the MCMC standard errors for the μ_is for patient 1 (a) and patient 2 (c); (Right column) Scatterplots of MCMC standard errors versus marginal posterior means of the μ_is for patient 1 (b) and patient 2 (d).

putationally efficient than simulating from the non-parametric Potts model proposed by Johnson et al. (2013). The MCMC algorithm for the CWAS model takes one hour to run on an iMac (3.2 GHZ Intel Core i5, 16 GB memory). The computational cost of the NP-Potts model is on the same order as that of the CWAS model only after the ratios of the normalizing constants have been computed. However, it takes more than one week to obtain good estimates of these ratios, rendering the algorithm impractical. Simulation results show that our model outperforms the two existing spatially adaptive smoothing models proposed by Brewer and Nolan (2007) and Reich and Hodges (2008). Finally, our model is easy to implement and converges faster than the other two.

Due to the computational intractability of the true likelihood, we resort to the pseudo-likelihood approximation proposed by Besag (1975) for the joint distribution of $\boldsymbol{\mu}$. Although the pseudo-likelihood approximation solves the problem and Besag (1977) showed that parameter estimation using this approach is efficient for simple Gaussian fields, using the pseudo-likelihood approximation brings up some issues. The MCMC samples do not come from the true model, and thus the inference is actually based on the likelihood from a misspecified model. It has been shown that the credible intervals based on MCMC samples from pseudo-likelihood have smaller empirical coverage than the nominal and less than the ones from true likelihood. Shaby (2014)

has proposed an adjustment for the MCMC samples, called the *open faced sandwich* (OFS) adjustment, to obtain the correct interval widths and coverage. We looked into the pseudo-likelihood adjustment method described by Shaby (2014) and found that it is impossible to apply it to our results. First, the OFS adjustment doesn't adjust the Bayesian estimators when the loss function used for Bayesian estimation is squared error loss. In this case, the Bayesian estimators are the posterior means and the OFS adjustment cancels out. Thus, the posterior means of the adjusted samples are exactly equal to that of the non-adjusted samples. Hence, under squared error loss, the OFS adjustment only adjusts the width of the credible intervals. In our case, that means that the OFS adjustment only affects the estimates of the m_is since the m_is depend on the variance of the MCMC samples. However, to do the adjustment, we would need to estimate $\Omega = Q^{-1}P^{\frac{1}{2}}Q^{\frac{1}{2}}$, where Ω, Q and P are $n \times n$ matrices, with n number of parameters. Although it is possible to estimate Q and P post hoc, our model has tens of thousands of parameters, thus to derive Ω we need to determine the inverse of a huge matrix, which is computationally intractable. We note that the examples considered by Shaby (2014) all have less than 5 parameters.

In fMRI studies, controlling the family-wise error rate (FWER) by restricting to anatomical region-of-interests (ROIs) can increase the statistical sensitivity since the number of multiple comparisons is drastically reduced. However, there is no loss (other than computational) for performing a whole brain analysis. In fact, a whole brain analysis has several advantages. In principle, brain activations are task- or condition-specific. This becomes obvious when you compare the activations in Figures 4 and 5. The tongue-twister task is phonetically and motorically challenging and therefore more centered to the anterior speech areas involving the precentral gyrus (Figure 4), while reading nonfinal embedded clause sentences involves more posterior speech and language areas including the angular gyrus and the so-called ventral stream of speech and language in general (Figure 5). However, contrary to the primary motor or visual system, for example, speech and language have no 'absolute' cortical representations (Bartsch et al., 2014). That is, you cannot use anatomical landmarks alone (like the central sulcus to delineate precentral gyrus of the primary motor system) to predict the areas where activations can be expected. In fact, this is reason why we often, and primarily, map speech and language. Furthermore, in presurgical fMRI of patients with brain lesions, compensatory mechanisms and 'neuroplasticity' may lead to atypical and unexpected activations (at the border or even outside anatomical ROIs defined by task-specific functional activations evoked in a group of healthy comparison subjects, for example) which we don't want to miss. Last but not least, performing a whole-brain analysis instead of an analysis restricted to predefined anatomical ROIs (as derived from prior functional studies, for example) or even just the intratumoral and peritumoral regions and the potential routes of neurosurgical access allows us to assess the general pattern of activation and deactivation, with both the expected as well as unexpected components (Figures 4 and 5), which proves very helpful to assure that the experiment worked (even in brain areas remote from the tumor). This is also why almost all presurgical fMRI studies record the data from the entire brain. Taken together, the whole-brain approach does not involve a statistical penalty for our analysis but provides us with a much richer set of information.

We applied the CWAS model to Z-statistics images from a mass univariate GLM approach. However, our model can be applied to the statistical output of independent component analyses (ICA) such as FSL's MELODIC (Beckmann and Smith, 2004). For speech and language mapping by fMRI, ICA may be more appropriate because it does not assume that brain activation is maintained at a similar level during one block (which is quite unlikely given the rather lengthy blocks of about 30 seconds in clinical fMRI).

A further extension to the CWAS model would be to modify the priors on the p_is so that they are explicitly spatially dependent a priori. In this paper, the prior distribution of the c_is is induced by the prior distribution on the p_is and the p_is/c_is are independent since $p_i \overset{iid}{\sim} \text{Beta}(\alpha, \beta), i = 1, \ldots, N$. If one wished to model the p_is as spatially dependent, while retaining control on the degree of smoothing, a prior which combines the feature of a CAR prior and a Beta distribution could be employed:

$$\pi(p_i) \propto \exp\left\{-0.5 \sum\nolimits_{i \sim j} (p_i - p_j)^2/\varphi^2 + (\alpha - 1)\log p_i + (\beta - 1)\log(1 - p_i)\right\}.$$

Another interesting extension of the CWAS model would be to modify it to accommodate multiple fMRI studies from the same patient, or multiple runs using the same paradigm. Consider, for example, a sentence completion experiment being performed in addition to a tongue-twister experiment. If there is any correlation between the two Z-statistic images, a bivariate CWAS model may be able to leverage it and produce more accurate results. Furthermore, instead of using it across different runs with different paradigms, such a model may especially be helpful in analyzing multiple runs using the same paradigm where, normally, a second level fixed-effects analysis is conducted. We are currently investigating several multivariate spatially adaptive CAR models for such analyses.

References

Bartsch, A. J., Biller, A., and Homola, G. (2014). "Presurgical tractography applications." In: Johansen-Berg, H. and Behrens, T. E. (eds.), *Diffusion MRI: from Quantitative Measurement to in-vivo Neuroanatomy*, 531–568. Elsevier Academic Press, Second edition. 18, 22

Bartsch, A. J., Homola, G., Biller, A., Solymosi, L., and Bendszus, M. (2006). "Diagnostic functional MRI: Illustrated clinical applications and decision-making." *Journal of Magnetic Resonance Imaging*, 23(6): 921–932. 2, 17

Beckmann, C. F. and Smith, S. M. (2004). "Probabilistic independent component analysis for functional magnetic resonance imaging." *IEEE Transactions on Medical Imaging*, 23(2): 137–152. 23

Besag, J. (1974). "Spatial interaction and the statistical analysis of lattice systems." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 36(2): 192–236. MR0373208. 3, 5

— (1975). "Statistical analysis of non-lattice data." *The Statistician*, 24(3): 179–195. 10, 21

— (1977). "Efficiency of pseudo-likelihood estimation for simple Gaussian fields." *Biometrika*, 64(3): 616–618. MR0494640. 10, 21

— (1993). "Towards Bayesian image analysis." *Journal of Applied Statistics*, 20(5–6): 107–119. 5

Besag, J., York, J., and Mollié, A. (1991). "Bayesian image restoration, with two applications in spatial statistics." *Annals of the Institute of Statistical Mathematics*, 43(1): 1–20. MR1105822. doi: http://dx.doi.org/10.1007/BF00116466. 3, 5

Bookheimer, S. (2007). "Pre-surgical language mapping with functional magnetic resonance imaging." *Neuropsychology Review*, 17(2): 145–155. 2

Brewer, M. J. and Nolan, A. J. (2007). "Variable smoothing in Bayesian intrinsic autoregressions." *Environmetrics*, 18(8): 841–857. MR2416406. doi: http://dx.doi.org/10.1002/env.844. 3, 6, 10, 21

Brook, D. (1964). "On the distinction between the conditional probability and the joint probability approaches in the specification of nearest-neighbour systems." *Biometrika*, 51(3/4): 481–483. MR0205315. 5, 9

Brooks, S. P. and Gelman, A. (1998). "General methods for monitoring convergence of iterative simulations." *Journal of Computational and Graphical Statistics*, 7(4): 434–455. MR1665662. doi: http://dx.doi.org/10.2307/1390675. 11, 18

Durnez, J., Moerkerke, B., Bartsch, A., and Nichols, T. E. (2013). "Alternative-based thresholding with application to presurgical fMRI." *Cognitive, Affective, & Behavioral Neuroscience*, 13(4): 703–713. 2

Fischl, B. (2012). "FreeSurfer." *Neuroimage*, 62(2): 774–781. 16

Flegal, J. M., Haran, M., and Jones, G. L. (2008). "Markov chain Monte Carlo: Can we trust the third significant figure?" *Statistical Science*, 250–260. MR2516823. doi: http://dx.doi.org/10.1214/08-STS257. 20

Flegal, J. M. and Hughes, J. (2012). *MCMCSE: Monte Carlo Standard Errors for MCMC*. Riverside, CA and Minneapolis, MN. R package version 1.0-1. 20

Friston, K., Jezzard, P., and Turner, R. (1994). "Analysis of functional MRI time-series." *Human Brain Mapping*, 1(2): 153–171. 2

Friston, K. J., Holmes, A. P., Poline, J., Grasby, P., Williams, S., Frackowiak, R. S., and Turner, R. (1995). "Analysis of fMRI time-series revisited." *Neuroimage*, 2(1): 45–53. 2

Fujiwara, N., Sakatani, K., Katayama, Y., Murata, Y., Hoshino, T., Fukaya, C., and Yamamoto, T. (2004). "Evoked-cerebral blood oxygenation changes in false-negative activations in BOLD contrast functional MRI of patients with brain tumors." *Neuroimage*, 21(4): 1464–1471. 17

Gelman, A. and Meng, X. L. (1998). "Simulating normalizing constants: From importance sampling to bridge sampling to path sampling." *Statistical Science*, 13(2): 163–185. MR1647507. doi: http://dx.doi.org/10.1214/ss/1028905934. 4

Gelman, A. and Rubin, D. B. (1992). "Inference from iterative simulation using multiple sequences." *Statistical Science*, 457–472. 11, 18

Gross, W. L. and Binder, J. R. (2014). "Alternative thresholding methods for fMRI data optimized for surgical planning." *Neuroimage*, 84: 554–561. 9

Haller, S. and Bartsch, A. J. (2009). "Pitfalls in fMRI." *Eur Radiology*, 19: 2689–2706. 4

Hastings, W. K. (1970). "Monte Carlo sampling methods using Markov chains and their applications." *Biometrika*, 57(1): 97–109. 10

Hickok, G. and Poeppel, D. (2007). "The cortical organization of speech processing." *Nature Reviews Neuroscience*, 8: 393–402. 14, 18

Holmes, A. P., Blair, R., Watson, G., and Ford, I. (1996). "Nonparametric analysis of statistic images from functional mapping experiments." *Journal of Cerebral Blood Flow & Metabolism*, 16(1): 7–22. 3

Hsu, Y. Y., Chang, C. N., Jung, S. M., Lim, K. E., Huang, J. C., Fang, S. Y., and Liu, H. L. (2004). "Blood oxygenation level-dependent MRI of cerebral gliomas during breath holding." *Journal of Magnetic Resonance Imaging*, 19: 160–167. 17

Johnson, T. D., Liu, Z., Bartsch, A. J., and Nichols, T. E. (2013). "A Bayesian non-parametric Potts model with application to pre-surgical FMRI data." *Statistical Methods in Medical Research*, 22(4): 364–381. MR3190664. doi: http://dx.doi.org/10.1177/0962280212448970. 4, 20, 21

Krüger, G. and Glover, G. H. (2001). "Physiological noise in oxygenation-sensitive magnetic resonance imaging." *Magnetic Resonance in Medicine*, 46(4): 631–637. 11

Kwong, K. K., Belliveau, J. W., Chesler, D. A., Goldberg, I. E., Weisskoff, R. M., Poncelet, B. P., Kennedy, D. N., Hoppel, B. E., Cohen, M. S., and Turner, R. (1992). "Dynamic magnetic resonance imaging of human brain activity during primary sensory stimulation." *Proceedings of the National Academy of Sciences*, 89(12): 5675–5679. 1

Lindquist, M. A. (2008). "The statistical analysis of fMRI data." *Statistical Science*, 23(4): 439–464. MR2530545. doi: http://dx.doi.org/10.1214/09-STS282. 2

Müller, P., Parmigiani, G., and Rice, K. (2007). "FDR and Bayesian multiple comparisons rules." In: Bernardo, J. M., Bayarri, M. J., Berger, J. O., Dawid, A. P., Heckerman, D., Smith, A. F. M., and West, M. (eds.), *Bayesian Statistics 8*, 349–370. Oxford Sci. Publ., Oxford Univ. Press, Oxford. MR2433200. 4, 8

Murray, I., Ghahramani, Z., and MacKay, D. (2006). "MCMC for doubly-intractable distributions." In: Dechter, R. and Richardson, T. S. (eds.), *Proceedings of the 22nd Annual Conference on Uncertainty in Artificial Intelligence (UAI)*, 359–366. Association for Uncertainty in Artificial Intelligence Press. 4

Nossek, E., Matot, I., Shahar, T., Barzilai, O., Rapoport, Y., Gonen, T., Sela, G., Grossman, R., Korn, A., Hayat, D., and Ram, Z. (2013). "Intraoperative seizures during awake craniotomy: incidence and consequences: analysis of 477 patients." *Neurosurgery*, 73(1): 135–140. 2

Ogawa, S., Lee, T., Kay, A., and Tank, D. (1990). "Brain magnetic resonance imaging with contrast dependent on blood oxygenation." *Proceedings of the National Academy of Sciences*, 87(24): 9868–9872. 1

Ogawa, S., Tank, D., Menon, R., Ellermann, J., Kim, S., Merkle, H., and Ugurbil, K. (1992). "Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging." *Proceedings of the National Academy of Sciences*, 89(13): 5951–5955. 3

Ojemann, J. G., Miller, J. W., and Silbergeld, D. L. (1996). "Preserved function in brain invaded by tumor." *Neurosurgery*, 39(2): 253–88. 18

Potts, R. B. (1952). "Some generalized order-disorder transformations." *Mathematical Proceedings of the Cambridge Philosophical Society*, 48: 106–109. MR0047571. 4

Reich, B. J. and Hodges, J. S. (2008). "Modeling longitudinal spatial periodontal data: A spatially adaptive model with tools for specifying priors and checking fit." *Biometrics*, 64(3): 790–799. MR2526629. doi: http://dx.doi.org/10.1111/j.1541-0420.2007.00956.x. 3, 6, 7, 10, 21

Shaby, B. A. (2014). "The open-faced sandwich adjustment for MCMC using estimating functions." *Journal of Computational and Graphical Statistics*, 23(3): 853–876. MR3224659. doi: http://dx.doi.org/10.1080/10618600.2013.842174. 21, 22

Smith, S. M., Jenkinson, M., Woolrich, M. W., Beckmann, C. F., Behrens, T. E., Johansen-Berg, H., Bannister, P. R., De Luca, M., Drobnjak, I., Flitney, D. E., Niazy, R. K., Saunders, J., Vickers, J., Zhang, Y., De Stefano, N., Brady, J. M., and Matthews, P. M. (2004). "Advances in functional and structural MR image analysis and implementation as FSL." *Neuroimage*, 23: S208–S219. 2, 11, 15

Sunaert, S. (2006). "Presurgical planning for tumor resectioning." *Journal of Magnetic Resonance Imaging*, 23(6): 887–905. 2

Ulmer, J. L., Hacein-Bey, L., Mathews, V. P., Mueller, W. M., DeYoe, E. A., Prost, R. W., Meyer, G. A., Krouwer, H. G., and Schmainda, K. M. (2004). "Lesion-induced pseudo-dominance at functional magnetic resonance imaging: implications for preoperative assessments, with discussion." *Neurosurgery*, 55: 569–591. 17

Welvaert, M., Durnez, J., Moerkerke, B., Verdoolaege, G., and Rosseel, Y. (2011). "neuRosim: an R package for generating fMRI data." *Journal of Statistical Software*, 44(10): 1–18. 11, 15

Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., and Smith, S. M. (2005). "Mixture models with adaptive spatial regularization for segmentation with an application to fMRI data." *IEEE Transactions on Medical Imaging,* 24(1): 1–11. 4

Woolrich, M. W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., and Smith, S. M. (2009). "Bayesian analysis of neuroimaging data in FSL." *Neuroimage*, 45(1): S173–S186. 2, 15

Worsley, K. J., Marrett, S., Neelin, P., Vandal, A. C., Friston, K. J., and Evans, A. C. (1996). "A unified statistical approach for determining significant signals in images of cerebral activation." *Human Brain Mapping*, 4(1): 58–73. 2

Yoo, S. S., Talos, I. F., Golby, A. J., Black, P. M., and Panych, L. P. (2004). "Evaluating requirements for spatial resolution of fMRI for neurosurgical planning." *Human Brain Mapping*, 21(1): 34–43. 3

Yue, Y., Loh, J. M., and Lindquist, M. A. (2010). "Adaptive spatial smoothing of fMRI images." *Statistics and Its Interface*, 3: 3–13. MR2609707. doi: http://dx.doi.org/10.4310/SII.2010.v3.n1.a1. 4

Acknowledgments

Dr. Johnson would like to thank the National Institute of Neurological Disorders and Stroke for their financial support: NIH Grant number 5-R01-NS-075066. The work presented in this manuscript represents the views of the authors and not necessarily that of the NINDS or NIH.