University of Michigan School of Public Health
The University of Michigan Department of Biostatistics Working Paper Series

Year 2007 *Paper* 75

A Bayesian Image Analysis of the Change in Tumor/Brain Contrast Uptake Induced by Radiation via Reversible Jump Markov Chain Monte Carlo

Xiaoxi Zhang* Tim Johnson†
Roderick J.A. Little‡

*University of Michigan, xiaoxi@umich.edu
†University of Michigan Biostatistics, tdjtdj@umich.edu
‡University of Michigan, rlittle@umich.edu

A Bayesian Image Analysis of the Change in Tumor/Brain Contrast Uptake Induced by Radiation via Reversible Jump Markov Chain Monte Carlo

Xiaoxi Zhang, Tim Johnson, and Roderick J.A. Little

Abstract

This work is motivated by a pilot study on the change in tumor/brain contrast uptake induced by radiation via quantitative Magnetic Resonance Imaging. The results inform the optimal timing of administering chemotherapy in the context of radiotherapy. A noticeable feature of the data is spatial heterogeneity. The tumor is physiologically and pathologically distinct from surrounding healthy tissue. Also, the tumor itself is usually highly heterogeneous. We employ a Gaussian Hidden Markov Random Field model that respects the above features. The model introduces a latent layer of discrete labels from an Markov Random Field (MRF) governed by a spatial regularization parameter. We further assume that conditional on the hidden labels, the observed data are independent and normally distributed, We treat the regularization parameter of the MRF, as well as the number of states of the MRF as parameters, and estimate them via the Reversible Jump Markov chain Monte Carlo algorithm. We propose a novel and nontrivial implementation of the Reversible Jump moves. The performance of the method is examined in both simulation studies and real data analysis. We report the pixel-wise posterior mean and standard deviation of the change in contrast uptake marginalized over the number of states and hidden labels. We also compare the performance with a Markov chain with fixed number of states and a parallel Expectation-Maximization approach from a frequentist perspective.

A Bayesian Image Analysis of the Change in Tumor/Brain Contrast Uptake Induced by Radiation via Reversible Jump Markov Chain Monte Carlo

Xiaoxi Zhang*, Timothy D. Johnson, and Roderick J. A. Little
Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109.

October 18, 2007

Abstract

This work is motivated by a pilot study on the change in tumor/brain contrast uptake induced by radiation via quantitative Magnetic Resonance Imaging. The results inform the optimal timing of administering chemotherapy in the context of radiotherapy. A noticeable feature of the data is spatial heterogeneity. The tumor is physiologically and pathologically distinct from surrounding healthy tissue. Also, the tumor itself is usually highly heterogeneous. We employ a Gaussian Hidden Markov Random Field model that respects the above features. The model introduces a latent layer of discrete labels from an Markov Random Field (MRF) governed by a spatial regularization parameter. We further assume that conditional on the hidden labels, the observed data are independent and normally distributed, We treat the regularization parameter of the MRF, as well as the number of states of the MRF as parameters, and estimate them via the Reversible Jump Markov chain Monte Carlo algorithm. We propose a novel and nontrivial implementation of the Reversible Jump moves. The performance of the method is examined in both simulation studies and real data analysis. We report the pixel-wise posterior mean and standard deviation of the change in contrast uptake marginalized over the number of states and hidden labels. We also compare the performance with a Markov chain with fixed number of states and a parallel Expectation-Maximization approach from a frequentist perspective.

Key Words: Quantitative MRI; Hidden Markov Random Fields; Reversible Jump MCMC, Trans-dimensional proposal, Swendsen-Wang Algorithm.

*Corresponding Author, email: xiaoxi@umich.edu

1 Introduction

As a non-invasive visualization tool, quantitative Magnetic Resonance Imaging (qMRI) enables researchers to assess the pathological and physiological alterations of living tissue, such as the brain. Recent work in measuring physiological changes (such as diffusion, perfusion, vascular permeability, and metabolism) in diseased tissue include Cao et al. (2005), Moffat et al. (2005), and Hamstra et al. (2005). It is expected that early prediction of (local) therapeutic efficacy during treatment via qMRI can lead to individualized treatment in the future.

Despite the advances in various treatment modes, the median survival time of high-grade glioma patients is about 1 year after diagnosis. This is largely attributable to the tight endothelia junction in the tumor (blood-tumor-barrier, BTB), which prohibits the passing of large chemotherapeutic agents and leads to low treatment efficacy. Although it is known that radiation can increase vascular permeability, recent attempts in sequential radiotherapy and chemotherapy have had limited success (Medical Research Council, 2001). This manuscript is motivated by a recent study on improving the effectiveness of chemotherapy among glioma patients via local radiation at the University of Michigan (Cao et al. 2005). This was the first study to use quantitative, high resolution MRI to assess the effectiveness of radiation on the tumor/brain vascular permeability of a molecule in the size range of chemotherapeutic agents used to treat high-grade gliomas (Cao et al. 2005). One of the goals was to determine the effects, over time, of radiation therapy on BTB relative to the blood-brain barrier (BBB). If it can be demonstrated that radiation therapy transiently increases the vascular permeability of the tumor to these large molecules, this suggests an optimal time for administering chemotherapy during radiation therapy, as opposed to waiting for the completion of radiation therapy.

Eleven volunteers with primary, high-grade gliomas were recruited in the study. They underwent T1-weighted qMRI with and without contrast enhancement prior to the initiation of radiation therapy. The same imaging protocol was subsequently performed after

approximately the first and third week of radiotherapy, and around 1, 3, and 6 months after the completion of radiotherapy. The contrast agent used, Gadolinium diethylenetriaminepentaacetic acid (Gd-DTPA), has approximately the same molecule diameter as many chemotherapeutic molecules; and hence its uptake rate can be used as a surrogate of tumor/brain vascular permeability to these drugs (Cao et al. 2005). In this work, we focus on the change in contrast uptake from baseline to week 3 (e.g. Figure 4a and 5a), which is of special interest to the investigator. The continuous observed data, denoted by $\mathbf{y}$, are represented as a range of gray levels.

Solid mass tumors are physiologically different from surrounding tissue, and the contrast uptake is highly heterogeneous. Thus the spatial continuity assumption usually made in fMRI analysis does not hold in this dataset. Many qMRI analyses ignore the inherent spatial correlation in the data (at the pixel level), and treat the data as independent observations (e.g. Cao et al. 2005; Moffat et al. 2005; Hamstra et al. 2005) which leads to incorrect variance estimates and invalid hypothesis tests.

To model the change of tumor/brain contrast uptake induced by radiation at the pixel-level, we use a model that accounts for the spatial correlation in the data and respects the distinct boundaries between tumor and healthy tissue. We introduce a layer of discrete hidden labels from a Markov Random Field (MRF, Besag 1974) which avoids over-smoothing of the data and accounts for spatial correlation. The MRF a priori encourages spatial continuity but allows for sapatial heterogeneity. Like many hidden Markov Models, we assume the observed data given the labels are independent and normally distributed.

We make the following additional comments on the proposed model: 1) The hidden labels are purely used to model the spatial correlation in the data, and are of not interest in and of themselves. We focus on the posterior distribution of change in contrast uptake. 2) The qMRI signal (contrast) at a single pixel measures the relaxation time of millions of nuclear spins averaged over hundreds of pulse sequences. Unlike positron emission tomography (PET) and single photon emission computed tomography (SPECT), there is no

mechanistically-based probabilistic models for MRI. However, Lei and Udupa (2002) investigate the statistical properties of MRI, and conclude that the normality assumption is reasonable. We also argue that in the MRI imaging process, spontaneous fluctuations (errors) from various sources add up, such as magnetic field inhomogeneities, thermal motion (Brownian motion) of free electrons in the electronics (mainly the radio frequency coil), and motion of the imaged object. Therefore, the central limit theorem provides sound ground for the normality assumption (see, e.g. , Liang and Lauterber, 1999, Ch. 8). The hidden MRF model, also known as the Potts model in statistical physics (Potts, 1952), has been applied in disease mapping (Green and Richardson, 2002), agriculture (Dryden, Scarr, and Taylor, 2003), and landscape genetics (Francois, Ancelet, and Guillot, 2006; Guillot, Estoup, Mortier, and Cosson, 2004). All cases share the same feature as our problem, spatial heterogeneity.

The spatial regularization parameter in MRF is traditionally treated as a fixed quantity (Green and Richardson, 2002; Francois, Ancelet, and Taylor, 2006). However, results are sensitive to the choice of this parameter (e.g., see Francois, Ancelet, and Guillot, 2006). Estimation of the spatial regularization parameter requires a corresponding normalizing constant, which is computationally intractable. We use the "thermal integration" method (Ogata 1989). Others adopt the pseudo-likelihood approach (Besag, 1974). As pointed out by Melas and Wilson (2002), this approach tends to overestimate the regularization parameter and over-smooth the data. Another prevalent practice is to assume the number of states of the MRF is known (Khalkighi, Soltanian-Zadeh, and Lucas, 2002; Marroquin, Arce, and Botello, 2003). This is reasonable in cases with substantial scientific justification, as when segmenting the brain into white matter, gray matter, and cerebrospinal fluid. However, the segmentation labels lack a strong biologically meaningful interpretation in our application. Therefore, it is natural to treat the number of labels as a model parameter, and estimate it using a Reversible Jump Markov chain Monte Carlo (RJMCMC) algorithm (Green, 1995; Richardson and Green, 1997). Zhang, Johnson, and Little (2007) employed the same model

in analysis of the same dataset and developed a maximum likelihood (ML) approach via the Expectation-Maximization (EM) algorithm with stochastic variation in the frequentist paradigm. In this manuscript, we take a complete Bayesian approach, and compare the performance of these two methods.

The current research builds on previous work by 1) estimating the spatial regularization parameter rather than assuming it known; 2) treating the number of states of MRF as a parameter and implementing RJMCMC on a large scale dataset; and 3) focusing on the posterior of local change in contrast uptake marginalized over the number of states and the hidden labels.

This manuscript is organized as follows. In the next section we introduce notation and specify the model. In section 3, we detail the implementation. In section 4, we first present results from simulation studies where we investigate the sensitivity to model assumption violations. We compare the results with the EM approach in Zhang, Johnson, and Little (2007). Results from the motivating example are also presented. We conclude by summarizing the strengths and weaknesses of our approach, and discussing future work.

2 Model Specification

We use the following notation. Pixels (short for picture element) will be indexed by $i = 1, 2, \cdots, N$. If pixel i and i' share a common edge, we call them neighbors, denoted $i \sim i'$. The set of neighbors of pixel i is denoted $\partial i = \{i' : i' \sim i\}$. The observed image is denoted by $\mathbf{y} = (y_1, y_2 \cdots, y_N)^{\mathrm{T}}$ (Figures 4a and 5a). In the proposed hidden MRF model, we introduce a latent discrete label Z_i from the state space $\mathcal{S} = \{1, 2, \cdots, M\}$ for each pixel i. The collection of latent labels, $\mathbf{z} = (Z_1 = z_1, Z_2 = z_2, \cdots, Z_N = z_N)^{\mathrm{T}}$, is called a configuration. The set of pixels with the same hidden label is referred to as a component, which can consist of disjoint clusters of pixels.

2.1 Distribution of the Data

We assume the observed data are conditionally independently distributed given the hidden labels, $(y_i \mid z_i = k, \mu_k, \sigma_k^2) \sim \mathrm{N}(\mu_k, \sigma_k^2)$, where $1 \leq k \leq M$. We write $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_M)^{\mathrm{T}}$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \cdots, \sigma_M^2)^{\mathrm{T}}$. The hidden labels follow a Gibbs distribution with probability mass function

$$\Pr(\mathbf{Z} = \mathbf{z} \mid \beta, M) = g^{-1}(\beta, M)\exp\Big\{\beta \sum_{i \sim i'} \mathrm{I}[z_i = z_{i'}]\Big\}, \tag{1}$$

where $\mathrm{I}[\cdot]$ is the indicator function. The regularization parameter β controls the spatial smoothness of MRF. When β is large, the correlation between pixels is strong (neighboring pixels have high tendency to assume the same label), and the configuration tends to be smooth. In principle, the MRF encourages neighboring pixels to assume the same label. Note that when $\beta = 0$, the pixels are independent. Furthermore, the spatial correlation decreases as the distance between pixels increases. The normalizing constant in Equation (1), $g(\beta, M) = \sum_{\mathbf{z} \in \mathcal{S}^N} \exp\Big\{\beta \sum_{i \sim i'} \mathrm{I}[z_i = z_{i'}]\Big\}$, has M^N outer summands, and therefore is not computationally tractable. We approximate it using the "thermal integration" method of Ogata (1989).

The joint distribution of the observed and unobserved data is

$$f(\mathbf{y}, \mathbf{Z} = \mathbf{z} \mid \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \beta, M) = \Pr(\mathbf{Z} = \mathbf{z} \mid \beta, M) \prod_{i=1}^{N} f(y_i \mid z_i, \mu_{z_i}, \sigma_{z_i}^2).$$

Although we assume conditional independence of the observed data given the hidden labels, one can show that the observed data are marginally correlated.

2.2 Prior Distributions

We specify

1. $\beta \sim \mathrm{Gamma}(\alpha_\beta, \beta_\beta)$;
2. $M \sim \mathrm{Poisson}(\lambda)$ truncated to $[M_{\min}, M_{\max}]$;

3. $\mu_k \sim \mathrm{N}(\mu_\mu, \sigma_\mu^2)$, for $k = 1, 2, ..., M$;

4. $\beta_\sigma \sim \mathrm{Gamma}(a, b)$;

5. $\sigma_k^2 \mid \beta_\sigma \sim \text{Inv-Gamma}(\alpha_\sigma, \beta_\sigma)$, for $k = 1, 2, ..., M$.

We fix $\alpha_\beta, \beta_\beta, \lambda, M_{\min}, M_{\max}, \mu_\mu, \mu_\sigma^2, a, b$, and α_σ as hyperpriors, the values of which are chosen in Section 2.3. We write all parameters $\boldsymbol{\theta} = (\boldsymbol{\mu}, \boldsymbol{\sigma}^2, \beta_\sigma, \beta, M)$, and assume conditional independence of the priors, i.e. $f(\boldsymbol{\theta}) = f(\beta)\Pr(M)f(\boldsymbol{\mu})f(\beta_\sigma)f(\boldsymbol{\sigma}^2 \mid \beta_\sigma)$.

Figure 1 illustrates the model structure in a Directed Acyclic Graph (DAG). Some earlier work (Chalmond, 1989; Won and Derin, 1992) avoid the normalizing constant in Equation (1) and use the pseudo-likelihood

$$\widetilde{\Pr}(\mathbf{Z} = \mathbf{z} \mid \beta, M) = \prod_{i=1}^{N} \frac{\exp\left\{\beta \sum_{i' \in \partial_i} \mathrm{I}[z_{i'} = z_i]\right\}}{\sum_{k=1}^{M} \exp\left\{\beta \sum_{i' \in \partial i} \mathrm{I}[z_{i'} = k]\right\}}$$

as an approximation of $\Pr(\mathbf{Z} = \mathbf{z} \mid \beta, M)$. Barker and Rayner (1997) show that under certain circumstances, this pseudo-likelihood may lead to an improper posterior distribution.

2.3 Hyperpriors

We follow the guideline given by Richardson and Green (1997). The prior mean (μ_μ) of the components is the midpoint of the observed range of data, while the prior variance (σ_μ^2) is half the squared range. This makes the prior distribution $\mathrm{N}(\mu_\mu, \sigma_\mu^2)$ fairly flat over the range of the data. We set $a = 2.5$ and $b = 5$ such that the prior mode of β_σ is 0.3, which favors a smaller component variance ($\alpha_\sigma = 2.1$). We choose $\lambda = 10$, $M_{\min} = 2$, and $M_{\max} = 20$, such that the number of components is neither too small nor too large. In Section 4.1.1 we investigate the sensitivity of the algorithm regarding this hyperprior. Furthermore, we set $\alpha_\beta = \beta_\beta = 2$.

2.4 Joint Distribution and Conditional Posterior Distributions

The kernel of the joint distribution function is

$$\begin{aligned} & \prod_{k=1}^{M} \left\{ {\sigma_\mu}^{-1} \exp\left[-0.5 {\sigma_\mu}^{-2} (\mu_k - \mu_\mu)^2 \right] \beta_\sigma^{\alpha_\sigma} (\sigma_k^{-2})^{\alpha_\sigma - 1} \exp(-\beta_\sigma \sigma_k^{-2}) \right\} \\ \times \quad & \prod_{i=1}^{N} \frac{1}{\sigma_{z_i}} \exp\left\{ -0.5 \sigma_{z_i}^{-2} (y_i - \mu_{z_i})^2 \right\} \times g^{-1}(\beta, M) \exp\Big\{ \beta \sum_{i \sim i'} \mathrm{I}[z_i = z_{i'}] \Big\} \\ \times \quad & \beta^{\alpha_\beta - 1} \exp(-\beta_\beta \beta) \times \beta_\sigma^{a-1} \exp(-b\beta_\sigma) \times \lambda^M (M!)^{-1} \times M!. \end{aligned} \tag{2}$$

We comment that Equation (2) is invariant to permutation of the components conditional on any M, which means the model parameters are not identifiable. In many fixed-dimension problems with segmentation as the main goal, the constraints $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k$ are imposed to resolve the identifiability issue. However, when M is treated as a parameter, the components can still be unidentifiable even with the constraints. We further discuss this towards the end of Section 3.3.2.

The conditional posterior distributions of $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$, β_σ, β, and $\mathbf{Z}$ are

$$(\mu_k \mid \cdot) \sim \mathrm{N}\left(\frac{\sum_{i \in D_k} y_i + \mu_\mu \sigma_k^2 / \sigma_\mu^2}{N_k + \sigma_k^2 / \sigma_\mu^2}, \left(\frac{N_k}{\sigma_k^2} + \frac{1}{\sigma_\mu^2} \right)^{-1} \right) \tag{3}$$

$$(\sigma_k^2 \mid \cdot) \sim \text{Inv-Gamma}\Big(0.5 N_k + \alpha_\sigma, 0.5 \sum_{i \in D_k} (y_i - \mu_k)^2 + \beta_\sigma \Big) \tag{4}$$

$$(\beta_\sigma \mid \cdot) \sim \mathrm{Gamma}\Big(a + M\alpha_\sigma, b + \sum_{k=1}^{M} \sigma_k^{-2} \Big) \tag{5}$$

$$f(\beta \mid \cdot) \propto g^{-1}(\beta, M) \exp\Big\{ \beta \sum_{i \sim i'} \mathrm{I}[z_i = z_{i'}] \Big\} \times \beta^{\alpha_\beta - 1} \exp(-\beta_\beta \beta) \tag{6}$$

$$\Pr(\mathbf{Z} = \mathbf{z} \mid \cdot) \propto \exp\Big\{ \beta \sum_{i \sim i'} \mathrm{I}[z_i = z_{i'}] \Big\} \prod_{i=1}^{N} \sigma_{z_i}^{-1} \exp\left\{ -0.5 \sigma_{z_i}^{-2} (y_i - \mu_{z_i})^2 \right\} \tag{7}$$

for $k = 1, 2, ..., M$, where $D_k = \{i : z_i = k\}$ denote components k and $N_k = |D_k|$ is the number of pixels in the component. Here the dot denotes all the rest of the parameters. We restrict the number of components M to change by no more than one at each iteration. The details of the trans-dimensional move are in Section 3.3.

2.5 Marginal Posterior Distribution of Change in Contrast Uptake

Our main goal is to establish the underlying change in contrast uptake, μ_{z_i}, and characterize it by its posterior mean, $\eta_i = \sum_{k=1}^{M} \mu_k \Pr(Z_i = k \mid \mathbf{y})$, and variance, $\psi_i^2 = \sum_{k=1}^{M} (\mu_k - \eta_i)^2 \Pr(Z_i = k \mid \mathbf{y})$. These quantities are estimated via Markov chain Monte Carlo (MCMC) draws, $\boldsymbol{\mu}^{(t)}$ and $\mathbf{z}^{(t)}$, at iteration t ($1 \leq t \leq T$), $\hat{\eta}_i = T^{-1} \sum_{t=1}^{T} \mu_{z_i^{(t)}}^{(t)}$ and $\hat{\psi}_i^{\,2} = T^{-1} \sum_{t=1}^{T} \left(\mu_{z_i^{(t)}}^{(t)}\right)^2 - \hat{\eta}_i^2$, where $\mu_{z_i^{(t)}}^{(t)} = \mu_k^{(t)}$ when $z_i^{(t)} = k$.

From an ML perspective, Zhang, Johnson, and Little (2007) proposed a similar measure, the "expected change in contrast uptake" calculated as $\sum_{k=1}^{M} \mu_k \Pr(Z_i = k \mid \mathbf{y}, \boldsymbol{\theta})$. However, this estimate depends on the Maximum Likelihood Estimate (MLE), $\hat{\boldsymbol{\theta}}$, and neglects the uncertainty in estimating it.

We also summarize the standard deviation of pixel intensity with its posterior mean, ϕ_i, and variance, ξ_i^2. The corresponding Monte Carlo estimates are $\hat{\phi}_i = T^{-1} \sum_{t=1}^{T} \sigma_{z_i^{(t)}}^{(t)}$ and $\hat{\xi}_i^{\,2} = T^{-1} \sum_{t=1}^{T} \left(\sigma_{z_i^{(t)}}^{(t)}\right)^2 - \hat{\phi}_i^2$, where $\sigma_{z_i^{(t)}}^{(t)} = \sigma_k^{(t)}$ when $z_i^{(t)} = k$.

3 Algorithm Details

The parameters $\boldsymbol{\mu}$, $\boldsymbol{\sigma}^2$, and β_σ can be updated via standard Gibbs sampling steps due to conjugacy (Equations (3), (4), and (5)). The spatial regularization parameter β requires a Metropolis-Hastings step. We update the hidden labels via the Swendsen-Wang algorithm (Swendsen and Wang, 1987), an efficient sampler for Potts models, as detailed in Section 3.2. The most difficult part of the sampler is updating the number of components. Since the dimension of the parameter space is determined by the number of components, we employ RJMCMC to update M.

3.1 Updating the Spatial Regularization Parameter

When updating the spatial regularization parameter at the t-th iteration, we use a Gaussian proposal distribution centered at the current value $\beta^{(t-1)}$, $\beta^* \sim \mathrm{N}(\beta^{(t-1)}, \sigma_{\text{pro}}^2)$, where the variance of the proposal, σ_{pro}^2, can be tuned. Since we use a symmetric proposal distribution,

the acceptance probability only depends on the ratio of the conditional posterior distributions in Equation (6). If the proposal is rejected, the current value is carried forward to the next iteration, $\beta^{(t)} = \beta^{(t-1)}$. We estimate the normalizing constant for a grid of β and M values ($\beta = 0, 0.05, \cdots, 3.00$ and $M = 1, 2, \cdots, 20$) via MCMC simulations off-line, and interpolate when necessary.

3.2 Updating the Hidden Labels via the Swendsen-Wang Algorithm

The conditional posterior distribution of the configuration in Equation (7) consists of two parts, an interaction term (between neighboring pixels) and a likelihood term (sometimes referred to as the external field, Higdon, 1998). A pixel-wise updating scheme does not mix well at the presence of the interaction term (Higdon, 1998). Hence, we use the Swendsen-Wang algorithm, an efficient sampling scheme designed to speed up the mixing of Potts models.

The Swendsen-Wang algorithm stochastically partitions the configuration into same-labeled contiguous regions such that the label for these regions can be updated independently. This clever idea is implemented in the following three steps.

1. We first generate a random variable, $u_{ii'} \sim \text{Uniform}(0, \exp\{\beta \text{I}[z_i = z_{i'}]\})$, for each pair of neighbors $i \sim i'$. Clearly, the joint distribution of all auxiliary variables $\mathbf{u} = \{u_{ii'}\}_{i \sim i'}$ is $f(\mathbf{u}) = \prod_{i \sim i'} \exp\{-\beta \text{I}[z_i = z_{i'}]\} \, \text{I}[0 \leq u_{ii'} \leq \exp\{\beta \text{I}[z_i = z_{i'}]\}]$.

2. By Bayes theorem, the posterior distribution of the labels conditional on the auxiliary variables is

$$\Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{u}, \boldsymbol{\theta}) \propto \prod_{i=1}^{N} \sigma_{z_i}^{-1} \exp\{-0.5 \sigma_{z_i}^{-2} (y_i - \mu_{z_i})^2\} \prod_{i \sim i'} \text{I}[0 \leq u_{ii'} \leq \exp\{\beta \text{I}[z_i = z_{i'}]\}]. \quad (8)$$

   The second term in Equation (8) defines the range over which the posterior distribution is non-zero. More concretely, when $1 \leq u_{ii'} \leq \exp\{\beta \text{I}[z_i = z_{i'}]\}$ holds, it builds a virtual stochastic "bond" between i and i' that requires both pixels to assume the same label.

That is if pixel i and i' assume the same label in the current configuration, they are "bonded" to assume a common label (could be different from the current one) with probability $1 - e^{-\beta}$. The bonds define an equivalence relation, and partition a configuration into contiguous regions of bonded pixels. We denote these equivalence classes by $C_1, C_2, ..., C_J$. According to the definition, a component consists of one or more such equivalence classes.

3. According to Equation (8), the labels of the equivalence classes can be updated independently. The new label $C_j^{(t+1)}$ of class j satisfies

$$\Pr(C_j^{(t+1)} = k \mid \mathbf{u}^{(t)}, \boldsymbol{\theta}^{(t)}) \propto \prod_{i \in C_j^{(t)}} \sigma_k^{-1(t)} \exp\left\{-.5\sigma_k^{-2(t)}\left(y_i - \mu_k^{(t)}\right)^2\right\}.$$

3.3 Updating the Number of Components via Reversible Jump MCMC

We estimate the number of components M via RJMCMC. Other applications of RJMCMC in Potts models include Green and Richardson (2002), who analyzed disease mapping data using a trans-dimensional proposal that emulates the spatial dependence structure of the Potts model; and Dryden, Scarr, and Taylor (2003), who applied a similar scheme to analyze weed and crop images. However, these proposal schemes only work for small or medium scale problems. With tens of thousands of pixels per image, as in our dataset, the acceptance rate is practically zero.

In the following, we propose a novel and nontrivial implementation of the trans-dimensional move inspired by the Swendsen-Wang algorithm. We first randomly choose between a split and a merge proposal with

$$\mathrm{P}_{\mathrm{split}}(M) = \begin{cases} 0 & \text{if } M = M_{\max} \\ 1 & \text{if } M = M_{\min} \\ 0.5 & \text{otherwise} \end{cases} \quad \text{and } \mathrm{P}_{\mathrm{merge}}(M) = \begin{cases} 0 & \text{if } M = M_{\min} \\ 1 & \text{if } M = M_{\max} \\ 0.5 & \text{otherwise} \end{cases}.$$

3.3.1 Split One Component into Two

If a split is chosen, we randomly pick a component k ($1 \le k \le M$) to split, i.e. $\mathrm{P}^{\mathrm{split}}_{\mathrm{select}}(k) = 1/M$. We need two extra parameters for the Gaussian mean and variance. To match the increase of dimension, we introduce two independent random variables, $u_1, u_2 \sim \mathrm{Beta}(2,2)$, and define a bijective transformation $(\mu^*_{k_1}, \mu^*_{k_2}, {\sigma^*_{k_1}}^2, {\sigma^*_{k_2}}^2) = \psi(\mu_k, \sigma^2_k, u_1, u_2)$ that matches the first two moments,

$$\begin{aligned}
\mu^*_{k_1} &= \mu_k - u_1\sigma_k, \\
\mu^*_{k_2} &= \mu_k + u_1\sigma_k, \\
{\sigma^*_{k_1}}^2 &= 2u_2(1-u_1^2)\sigma^2_k, \\
{\sigma^*_{k_2}}^2 &= 2(1-u_2)(1-u_1^2)\sigma^2_k.
\end{aligned}$$

We denote the new set of parameters with superscript *, $\boldsymbol{\mu}^* = (\mu^*_1, \mu^*_2, \cdots, \mu^*_{M+1})^{\mathrm{T}}$ and $\boldsymbol{\sigma}^{*2} = (\sigma^{*2}_1, \sigma^{*2}_2, \cdots, \sigma^{*2}_{M+1})^{\mathrm{T}}$. It is intuitive that when there exists a component h ($1 \le h \le M+1$) satisfying $\mu^*_{k1} < \mu^*_h < \mu^*_{k2}$, the proposal is not likely to be accepted. We therefore reject such proposals.

We propose a new labeling for each contiguous equivalence class defined by common hidden labels. In contrast to the Swendsen-Wang update in Section 3.2 which further breaks contiguous same-labeled regions with stochastic "bonds", here we deterministically partition the configuration. This is equivalent to a Swendsen-Wang update with an infinite regularization parameter ($\beta \to \infty$) where all same-labeled neighbors are "bonded".

Suppose there are L such classes $\xi_1, \xi_2, \cdots, \xi_L$ with current labels $E_1, E_2, \cdots, E_L$ (by definition, $L \ge M$, and when $i \in \xi_l$, $z_i = E_l$). In our split proposal, we randomly draw the new label of these equivalence classes with probability proportional to the likelihood. Specifically, the probability of class ξ_l assuming a new label k is

$$p^l_k \propto \prod_{i \in \xi_l} \sigma^{*-1}_k \exp\left\{-0.5\sigma^{*-2}_k (y_i - \mu^*_k)^2\right\}, \text{for } k = 1, 2, \cdots, M+1, \tag{9}$$

where the p^l_k are normalized so that $\sum_{k=1}^{M+1} p^l_k = 1$ for all l. Let E^*_l denote the new label of

class ξ_l from Equation (9), i.e. $p^l_{E^*_l} = \Pr(z^*_i = E^*_l, \forall i \in \xi_l)$. The allocation probability in the current state conditional on the equivalence classes is $\mathrm{P}_{\mathrm{alloc}}(\mathbf{z}) = \prod_{l=1}^{L} p^l_{E^*_l}$.

Note that in the above proposal, any two adjacent equivalence classes must assume different labels in the current configuration by definition. However, the two classes can possibly assume the same new label in the proposed configuration. Hence, the resulting new configuration appears "smoother" with fewer different-labeled neighbors. The sum of the interaction terms $\sum_{i\sim i'} \mathrm{I}[z_i = z_{z'_i}]$ is guaranteed nondecreasing.

We also compute the allocation probability for the new configuration given the same equivalence classes, $\mathrm{P}_{\mathrm{alloc}}(\mathbf{z}^*) = \prod_{l=1}^{L} p^{l^*}_{E_l}$, where $p^{l^*}_k \propto \prod_{i\in\xi_l} \sigma_k^{-1} \exp\left\{-0.5\sigma_k^{-2}(y_i - \mu_k)^2\right\}$, for $k = 1, 2, \cdots, M$.

The overall acceptance probability for a split proposal is $\min(1, A)$ with

$$\begin{aligned} A &= \frac{\Pr(\mathbf{z}^*, \boldsymbol{\mu}^*, \boldsymbol{\sigma}^{2*}, \beta_\sigma, \beta, M+1 \mid \mathbf{y})}{\Pr(\mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2, \beta_\sigma, \beta, M \mid \mathbf{y})} \times \frac{\mathrm{P}_{\mathrm{merge}}(M+1)\mathrm{P}^{\mathrm{merge}}_{\mathrm{select}}(k_1, k_2)\mathrm{P}_{\mathrm{alloc}}(\mathbf{z}^*)}{\mathrm{P}_{\mathrm{split}}(M)\mathrm{P}^{\mathrm{split}}_{\mathrm{select}}(k)\mathrm{P}_{\mathrm{alloc}}(\mathbf{z})b(u_1)b(u_2)} \\ &\quad \times \left| \frac{\partial(\mu^*_{k_1}, \mu^*_{k_2}, {\sigma^*_{k_1}}^2, {\sigma^*_{k_2}}^2)}{\partial(\mu_k, \sigma^2_k, u_1, u_2)} \right|. \end{aligned} \tag{10}$$

Here $b(\cdot)$ is the density function of the Beta$(2, 2)$ distribution. The first line is the product of the posterior ratio and the proposal ratio, while the second line is the Jacobian of the bijective transformation.

3.3.2 Merge Two Components into One

The merge proposal is determined by the above split move. We randomly pick a pair of components with adjacent means k_1 and k_2 ($\mathrm{P}^{\mathrm{merge}}_{\mathrm{select}}(k_1, k_2) = 1/(M-1)$). Reversing the bijective transformation in the split proposal, we compute the parameters for the new component by

$$\begin{aligned} \mu^*_k &= 0.5\,(\mu_{k_1} + \mu_{k_2})\,, \\ {\mu^*_k}^2 + {\sigma^*_k}^2 &= 0.5(\mu^2_{k_1} + \sigma^2_{k_1} + \mu^2_{k_2} + \sigma^2_{k_2}). \end{aligned}$$

The new set of parameters are $\boldsymbol{\mu}^* = (\mu^*_1, \mu^*_2, \cdots, \mu^*_{M-1})^{\mathrm{T}}$ and $\boldsymbol{\sigma}^{*2} = (\sigma^{*2}_1, \sigma^{*2}_2, \cdots, \sigma^{*2}_{M-1})^{\mathrm{T}}$.

Again, we propose a new configuration $\mathbf{z}^*$ with $M-1$ components based on the same-labeled contiguous equivalence classes. The new label E^*_l of class ξ_l is drawn from $p^l_k \propto$

$\prod_{i\in\xi_l}\sigma_k^{*-1}\exp\left\{-0.5\sigma_k^{*-2}(y_i-\mu_k^*)^2\right\}$, for $k=1,2,\cdots,M-1$. The associated allocation probability is $\mathrm{P}_{\mathrm{alloc}}(\mathbf{z})=\prod_{l=1}^{L}p_{E_l^*}^l$. Similarly, the allocation probability for the new configuration is $\mathrm{P}_{\mathrm{alloc}}(\mathbf{z}^*)=\prod_{l=1}^{L}p_{E_l}^{l^*}$, where $p_k^{l^*}\propto\prod_{i\in\xi_l}\sigma_k^{-1}\exp\left\{-0.5\sigma_k^{-2}(y_i-\mu_k)^2\right\}$, for $k=1,2,\cdots,M$.

Similar to Equation (10), the overall acceptance probability is $\min(1,B)$, where

$$B = \frac{\Pr(\mathbf{z}^*,\boldsymbol{\mu}^*,\boldsymbol{\sigma}^{2*},\beta_\sigma,\beta,M-1\mid\mathbf{y})}{\Pr(\mathbf{z},\boldsymbol{\mu},\boldsymbol{\sigma}^2,\beta_\sigma,\beta,M\mid\mathbf{y})}\times\frac{\mathrm{P}_{\mathrm{split}}(M-1)\mathrm{P}_{\mathrm{select}}^{\mathrm{split}}(k)\mathrm{P}_{\mathrm{alloc}}(\mathbf{z}^*)b(u_1)b(u_2)}{\mathrm{P}_{\mathrm{merge}}(M)\mathrm{P}_{\mathrm{select}}^{\mathrm{merge}}(k_1,k_2)\mathrm{P}_{\mathrm{alloc}}(\mathbf{z})}\times\left|\frac{\partial(\mu_k^*,\sigma_k^{*2},u_1,u_2)}{\partial(\mu_{k_1},\mu_{k_2},\sigma_{k_1}^2,\sigma_{k_2}^2)}\right|.$$

We mentioned in Section 2.3 that the identifiability issue exists in RJ even under the constraints ($\mu_1\le\mu_2\le\cdots\le\mu_M$). For instance, suppose there are three components labeled 1, 2, and 3, after the merge of component 1 and 2 is accepted, the component with original label 3 is automatically relabeled as 2 under the constraints. It is clear that the component with new label 2 (originally labeled 3) is not the component with original label 2. Fortunately, we are only interested in the "true" change in contrast uptake, and will marginalize over M as well as the labels as discussed in Section 2.5. The identifiability of the parameters is not of our main concern. We employ the constraints to automatically reject unlikely split proposals and to propose two components to merge.

4 Results

To evaluate the performance of the proposed method, we conduct a set of simulation studies. We compute the Mean Squared Error (MSE) of pixel mean ($\mathrm{MSE}(\mu_z)=N^{-1}\sum_{i=1}^{N}(\hat{\eta}_i-\mu_{z_i}^{\mathrm{true}})^2$) and standard deviation ($\mathrm{MSE}(\sigma_z)=N^{-1}\sum_{i=1}^{N}(\hat{\psi}_i-\sigma_{z_i}^{\mathrm{true}})^2$), where $\mu_{z_i}^{\mathrm{true}}$ and $\sigma_{z_i}^{\mathrm{true}}$ are the "truth". We run multiple simulations and compute the average MSE and bias. We also implement the EM approach discussed by Zhang, Johnson, and Little (2007). The performance of the two approaches is compared under the conditionally independent noise assumption, as well as under various violations of the assumption. Analysis of a real dataset is also presented.

4.1 Simulation Study with Conditionally Independent Noise

We first divide a 128×128 lattice into 64 blocks of size 16×16. We assume the observed intensities of pixels within the same block follow the same distribution, which is randomly chosen from eight candidate Gaussian distributions ($\boldsymbol{\mu} = (-5.95, -4.25, -2.55, -0.85, 0.85, 2.55, 4.25, 5.95)^{\mathrm{T}}$, $\boldsymbol{\sigma} = (1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0, 1.0)^{\mathrm{T}}$). An example of pixel-wise true mean ($\mu_{z_i}^{\mathrm{true}}$) and observed data ($y_i$) is found in Figures 2. We simulate twenty sets of such images.

Zhang, Johnson, and Little (2007) report biased parameter estimates when the spatial structure is ignored (i.e. fixing $\beta = 0$), even if the number of components is held at the true value ($M = 8$). This is essentially a Gaussian mixture model (GMM) with equal component weights. We experience similar results; the average MSEs are $\overline{\mathrm{MSE}}_{\mathrm{GMM}}^{M=8}(\mu_z) = 0.849 \pm 0.043$ (mean $\pm$ standard deviation) and $\overline{\mathrm{MSE}}_{\mathrm{GMM}}^{M=8}(\sigma_z) = 0.032 \pm 0.021$. We then run the EM algorithm with stochastic variation discussed by Zhang, Johnson, and Little (2007) with M fixed at the true value ($M = 8$). The initial values of $\boldsymbol{\mu}$ are evenly spaced on the range of the data. The MSEs are lower than for GMM, namely, $\overline{\mathrm{MSE}}_{\mathrm{EM}}^{M=8}(\mu_z) = 0.119 \pm 0.085$ and $\overline{\mathrm{MSE}}_{\mathrm{EM}}^{M=8}(\sigma_z) = 0.010 \pm 0.009$.

We also run a fixed-dimension MCMC ($M = 8$) with β as a parameter. The average MSEs are smaller than the EM approach, $\overline{\mathrm{MSE}}_{\mathrm{MCMC}}^{M=8}(\mu_z) = 0.040 \pm 0.028$ and $\overline{\mathrm{MSE}}_{\mathrm{MCMC}}^{M=8}(\sigma_z) = 0.0002 \pm 0.003$. We then run the proposed algorithm for 20000 iterations after 20000 burn-in iterations on each simulation set. The RJ acceptance rate ranges from 4% to 20%. We display the posterior mean ($\hat{\eta}_i$) and standard deviation ($\hat{\psi}_i$) for one such simulation in Figure 2. The proposed algorithm further reduces the MSEs, $\overline{\mathrm{MSE}}_{\mathrm{RJ}}(\mu_z) = 0.028 \pm 0.003$ and $\overline{\mathrm{MSE}}_{\mathrm{RJ}}(\sigma_z) = 1 \times 10^{-4} \pm 3 \times 10^{-4}$. All three approaches have minimal bias in estimating pixel mean, $-2 \times 10^{-4} \pm 0.008$ for EM, $-4 \times 10^{-4} \pm 0.007$ for fixed-dimension MCMC, and $-4 \times 10^{-4} \pm 0.007$ for RJ.

We reason that the relatively poor performance of the EM approach is due to the fact that about half of the EM runs are trapped at local modes. The fixed-dimension MCMC explores the parameter space better. The extra flexibility in estimating the number of states

in RJ further improves the mixing of the Markov chain, and results in even smaller MSEs.

4.1.1 Sensitivity Analysis

Green (1995) reports that the RJMCMC algorithm is sensitive to the choice of hyperprior for the number of states λ. We run the proposed algorithm with $\lambda = 5$ and 15 in addition to the original value 10 chosen in Section 2.3. In our case, the marginal posterior mean of pixel intensities are almost identical under different λ (top row of Figure 3). The posterior distributions are not highly sensitive to the choice of the prior parameter. The average MSE of pixel mean intensity under all three scenarios are all 0.028 ± 0.003.

We also examine the proposed method under various signal to noise ratios. The current set of simulation studies has signal level $\Delta_\mu = \mu_{k+1} - \mu_k = 1.7$ and noise level $\sigma = 1.0$. A significant portion of posterior draws of M lies at the true value 8, which suggests the data contain considerable amount of information on the number of components. In other simulations, we varied the signal level ($\Delta_\mu = 1.6, 1.5, \cdots, 1.0$) with the same noise level. The average MSE increases as the signal level decreases from 0.031 ± 0.007 for $\Delta_\mu = 1.6$, 0.049 ± 0.026, 0.065 ± 0.035, 0.121 ± 0.043, 0.157 ± 0.058, 0.159 ± 0.030, to 0.192 ± 0.058 for $\Delta_\mu = 1.0$. We observe M as low as 4 with a significant portion of posterior density for M less than 8.

4.2 Simulation Studies with Correlated Noise

We apply Gaussian smoothing kernels on the simulated data to evaluate the robustness of our method to violations of the conditional independence assumption. We follow the same Gaussian kernel parameters as Zhang, Johnson, and Little (2007) ($\sigma = 0.42, 0.85, 1.70$, and 3.40). Note that large σ indicates heavy smoothing.

We display the smoothed observed image and the results for $\sigma = 0.85$ and 3.40 in the bottom two rows of Figure 2. The "edge preservation" of the proposed algorithm is more evident when the degree of smoothing is small. The average MSEs under all degree of smoothing are displayed in Table 1. The average MSE of pixel intensity of the proposed

method is uniformly lower for the RJMCMC approach than for the EM approach under all degrees of smoothing. This occurs despite the fact that the number of components is fixed at the "true" value in the EM approach. The difference in average MSE decreases as the smoothing parameter σ increases.

4.3 Application

In the motivating study, eleven subjects received fractionated three-dimensional conformal radiation with a median dose of 70 Gy at 2 Gy per fraction, and underwent Gd-DTPA contrast enhanced T1-weighted qMRI before, during, and after treatment. All images were registered to anatomical Computed Tomography (CT) images obtained for treatment planning purpose. The natural logarithm of the ratio of the post- and pre-enhanced T1-weighted qMRI images are used as the Gd-DTPA contrast uptake index after image normalization (Cao et al. 2005). We apply the proposed algorithm on a subset of the data that is of special research interest, i.e. the pre-radiation visit and the visit at approximately 3 weeks into radiotherapy. Due to space limitations, we only display the images on the same two subjects as reported by Zhang, Johnson, and Little (2007). The results on all subjects are reported in Table 2.

We run the proposed algorithm for 50000 iterations after 25000 burn-in iterations. The acceptance rate of the trans-dimensional proposals is 12.5% for subject 1. The marginal posterior mean of change in contrast uptake and associated standard deviation (Figure 4b and 4c) delineate the heterogeneous response of the tumor. The results from subject 2 are similar (Figure 5b and 5c).

4.3.1 Interpretation of the Results

In the original analysis, Cao et al. (2005) divided the tumor into two regions, one with relatively high pre-treatment contrast uptake and the other with relatively low initial contrast uptake. This more-or-less divides the tumor into a "core" (low initial contrast uptake) surrounded by an "annulus" (high initial contrast uptake). It can be argued that the core of

the tumor is typically hypoxic (low oxygen content) due to a lack of blood supply, which is known to have a protective effect against damage due to both radiation and chemotherapy. This region is usually a source of tumor regrowth. Hence, the focus was on demonstrating that radiation therapy has a transient effect on the core of the tumor with respect to increasing the contrast uptake. The time when this increase is greatest can be an optimal time for initiating chemotherapy, allowing for more effective control of the tumor core. As in Zhang, Johnson, and Little (2007), we smooth the baseline contrast uptake image under the spatial model, and use the 95^{th} percentile of the healthy tissue contrast uptake as a cutoff to divide the tumor into "core" and "annulus".

As discussed in the introduction, large increases in contrast uptake are indicative of heavier damage to the BTB/BBB. This suggests that chemotherapeutic agents, in the size range of the contrast medium, may pass the BTB/BBB more easily. Hence, a large increase in the tumor and a small increase (or even decrease) in healthy tissue may suggest the opportunity to deliver these agents more effectively during this window of time. An alternative to comparing the mean change is to define a threshold of change and compare the proportions of healthy and diseased tissue that exceed this threshold. Without an established biologically meaningful threshold of change in this exploratory study, we first apply the illustrative threshold, 0.06, used by Zhang, Johnson, and Little (2007).

For subject 1, 8.2% of the normal tissue exceeds the threshold, while 59.5% of the tumor "core" has an increase in contrast uptake above the threshold, compared to 15.6% in the tumor "annulus" (Figure 4e). Similarly, for subject 2, 4.0% of the healthy tissue exceeds the threshold. While 54.3% of the tumor "core" has a change in uptake that exceeds the threshold, compared to 12.7% in the "annulus" (Figure 5e). The percentages are very similar to those reported by Zhang, Johnson, and Little (2007).

We take the threshold idea one step further, and argue that the Area Under the Curve (AUC) of a Receiver Operating Characteristic (ROC) curve can be used to better summarize the the differential response of the tumor "core/annulus". Although the AUC is

traditionally used to quantitate a diagnostic test, it essentially provides a distribution-free summary statistic of the distance between two distributions (Pepe, 2003), namely the tumor (core/annulus) and the healthy tissue in our case. The AUC can be interpreted as the probability of a randomly picked tumor pixel having a higher mean change in contrast uptake than a randomly picked normal pixel. A larger AUC value, which indicates better separation between the tumor and the brain, is desirable. The AUC is computed at each iteration for both the tumor "core" and "annulus" versus the healthy tissue.

The 95% posterior credible interval of AUC between the tumor "core" and the healthy tissue is $(0.74, 0.76)$ for subject 1, which suggests the "core" region has a substantial increase on average in contrast uptake compared to the healthy tissue. The tumor "core" of subject 2 is also well separated from the healthy tissue with 95% posterior credible interval $(0.81, 0.82)$. The AUC between the tumor "annulus" and the healthy tissue is under 0.5 for both subjects ($(0.23, 0.25)$ for subject 1 and $(0.25, 0.26)$ for subject 2), which means the "annulus" on average has decreased contrast uptake compared to the healthy tissue. The AUC for all eleven subjects are given in Table 2. Most subjects, except for one, demonstrate significantly increased contrast uptake in the tumor "core" relative to the healthy tissue.

4.3.2 Sensitivity Analysis

We report the marginal posterior distribution of M under various choices of prior parameters ($\lambda = 5$, 10, and 15) in the bottom two rows of Figure 3. Again, the posterior distribution of M as well as the Monte Carlo estimate $\hat{\eta}_i$ are relatively unaffected by the choice of λ.

Admittedly, the choice of 95^{th} percentile for defining the tumor "core" and "annulus" is ad-hoc. We performed a small study to address the sensitivity of our results to this choice. We used alternative cutoffs of the 90^{th} and 97.5^{th} percentiles of the healthy tissue contrast uptake. The percentages of the "core" and "annulus" of the tumors that exceed the various thresholds are given in Table 2. Although the AUC varies as the threshold of "core/annulus" changes, the trend is the same.

5 Conclusion

In this article, we implement a statistical imaging model that respects the key feature of spatial heterogeneity in the qMRI data. Compared to previous work, we integrate many parts that have been previously discussed individually. First, we estimate the normalizing constant via thermal integration, instead of using the pseudo-likelihood. Second, we estimate the spatial regularization parameter from the data, rather than holding it fixed. Third, we acknowledge that there is no clear substantive knowledge regarding the number of components, given the lack of physical interpretation of the hidden labels. We therefore marginalize over both, and focus on the marginal posterior distribution of the pixel change in contrast uptake, which is of primary scientific interest. We also note that the implementation of the RJ steps is non-trivial. The split scheme proposed by Richardson and Green (1997) is inadequate for our application, although it has proved successful in smaller scale problems. The split proposal we use is inspired by the Swendsen-Wang algorithm, and works well because it produces smoother new configurations. Finally, we propose the use of AUC to summarize the differential change in contrast uptake of the tumor "core/annulus" versus the healthy tissue.

In some respects, the current work is a Bayesian parallel of the EM approach proposed by Zhang, Johnson, and Little (2007). However, we find in the simulation studies that the proposed method is less sensitive to the initial values chosen. Also, the RJ approach yields estimates with smaller MSE than the fixed-dimension MCMC, and the fixed-dimension MCMC yields smaller MSE than the EM approach. The comparative advantage is greatest when the data are analyzed without any kernel smoothing. We also find that when the data-generation mechanism is known, a fixed-dimension MCMC performs comparably with the proposed method. However, the loss associated with mis-specified mechanism can be severe. We therefore conclude the proposed method robust and efficient.

An alternative MCMC technique dealing with variable and unknown parameter space dimension is the birth-and-death MCMC (BDMCMC, Stephens 2000). Extending some

early work, BDMCMC shares the essence of RJMCMC in variable dimensional problems, and can be further generalized to continuous-time MCMC. It is shown in Cappe, Robert, and Ryden (2003) section 4 that under some week regularity conditions, there exists a sequence of RJMCMC processes converging to a birth-and-death MCMC sampler. In the nonparametric Bayesian field, there is a large body of literature on infinite mixture models using the Dirichlet prior. Another possible extension is an infinite spatial mixture model (Guillot, Estoup, Mortier, and Cosson, 2005, Francois, Ancelet, and Taylor, 2006). However, implementation of both alternatives may be challenging given the large volume of data.

It appears that model fitting is improved by using RJMCMC than EM. However, we comment that the image model itself is insufficient to provide an integrated picture of the differential profiles of the tumor/brain contrast uptake induced by radiation. The model only utilizes a subset of the data (the baseline and week 3) that is of primary research interest. In a sense, it emulates a hypothesis testing scenario, in which the credibility of a null hypothesis is examined. Obviously, this only provides partial information about the possible optimal timing of administering chemotherapy with respect to radiation. To capture the entire profile of contrast uptake, we need a model that simultaneously captures the spatial and temporal correlation inherent in the data. This is a future direction of our research.

Acknowledgments

The authors would like to thank Yue Cao, Department of Radiation Oncology, University of Michigan for providing us with the data. This work was partially funded by NIH grants PO1 CA087684-5 and PO1 CA59827-11A1.

References

[1] Medical Research Council (2001), "Randomized Trial of Procarbazine, Lomustine, and Vincristine in the Adjuvant Treatment of High-Grade Astrocytoma: A Medical Research Council Trial", *J. Clin. Oncol.*, 19(2):509-518.

[2] Barker and Rayner, Unsupervised Image Segmentation Using Markov Random Field Models, 1997.

[3] Besag J (1974), "Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussions)", *Journal of the Royal Statistical Society. Series B*, 36:192-236.

[4] Chalmond B (1989), An Iterative Gibbsian Technique for Reconstruction of M-ary Images, *Pattern Recognition* **22**, 747-761.

[5] Cao Y, Tsien CI, Shen Z, Tatro DS, Ten Haken R, Kessler ML, Chenevert TL, and Lawrence TS (2005), "Use of Magnetic Resonance Imaging to Assess Blood-Brain/Blood-Glioma Barrier Opening During Conformal Radiotherapy", *Journal of Clinical Oncology*, 23(18): 4127-4136.

[6] Dryden IL, Scarr MR, and Taylor CC (2003), Bayesian Texture Segmentation of Weed and Crop Images Using Reversible Jump Markov Chain Monte Carlo Methods, *Journal of Royal Statistical Society, Series C* **52**, 31-50.

[7] Francois O, Ancelet S, and Guillot G (2006), "Bayesian Clustering Using Hidden Markov Random Fields in Spatial Population Genetics", *Genetics*, 174:805-816.

[8] Green PJ (1995), "Reversible Jump Markov Chain Monte Carlo Computation and Bayesian Model Determination", *Biometrika*, 82(4):711-732.

[9] Richardson S and Green PJ (1997), "On Bayesian Analysis of Mixtures with an Unknown Number of Components", *Journal of the Royal Statistical Society. Series B*, 59(4):731-792.

[10] Green PJ and Richardson S (2002), "Hidden Markov Models and Disease Mapping", *Journal of the American Statistical Association*, 97:1055-1070.

[11] Guillot G, Estoup A, Mortier F, and Cosson JF (2005), "A Spatial Statistical Model for Landscape Genetics", *Genetics*, 170(3):1261-80.

[12] Hamstra DA, Chenevert TL, Moffat BA, Johnson TD, Meyer CR, Mukherji SK, Quint

DJ, Gebarski SS, Fan X, Tsien CI, Lawrence TS, Junck L, Rehemtulla A, and Ross BD (2005), "Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma", *PNSA*, 102(46):16759-16764.

[13] Higdon D (1998), "Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications", *Journal of the American Statistical Association*, 93.

[14] Khalighi MM, Soltanian-Zadeh HS, and Lucas C (2002), Unsupervised MRI Segmentation with Spatial Connectivity, *P*roceedings of SPIE, Medical Imaging **4684**, 1742-1750.

[15] Liang ZP and Lauterbur PC (1999), Principles of Magnetic Resonance Imaging: A Signal Processing Perspective, SPIE press book, Chapter 8 section 2.

[16] Lei T and Udupa JK (2002). Statistical Properties of X-ray CT and MRI from Imaging Physics to Image Statistics, *Proceedings of the SPIE, Medical Imaging* **4682**, 82-93.

[17] Melas DE and Wilson SP (2002), Double Markov Random Fields and Bayesian Image Segmentation, *Institute of Electrical and Electronics Engineers Transactions on Signal Processing* **50**, 357-365.

[18] Moffat BA, Chenevert TL, Lawrence TS, Meyer CR, Johnson TD, Dong Q, Tsien C, Mukherji S, Quint DJ, Gebarski SS, Robertson PL, Junck LR, Rehemtulla A, and Ross BD (2005), "Functional diffusion map: A noninvasive MRI biomarker for early stratification of clinical brain tumor response", *PNSA*, 102(15):5524-5529.

[19] Marroquin JL, Arce E, and Botello S (2003), Hidden Markov Measure Field Models for Image Segmentation, *I*EEE Transactions on Pattern Analysis and Machine Intelligence **25**, 1380-1387.

[20] Pepe MS (2003), "The Statistical Evaluation of Medical Tests for Classification and Prediction", *Oxford Press*, Chapter 4.3.

[21] Potts RB (1952), Some generalized order-disorder transformations, *Proceedings of the Cambridge Philosophical Society* **48**, 106-109;

[22] Stephens M (2000), Bayesian Analysis of Mixture Models with an Unknown Number of Components - An Alternative to Reversible Jump Methods, *The Annals of Statistics* **28**, 40-74.

[23] Won CS and Derin H (1992), Unsupervised Segmentation of Noisy and Textured Images Using Markov Random Fields, *Graphical Models and Image Processing* **54**, 308-328.

[24] Zhang X, Johnson TD, and Little RJA (2007), Quantitative Magnetic Resonance Image Analysis via the EM Algorithm with Stochastic Variation, *The University of Michigan Department of Biostatistics Working Paper Series*, Working Paper **72**.

Table 1: Average MSE of the proposed method vs. the ML approach

Smoothing	Reversible Jump		Maximum Likelihood	
Kernel (σ)	$\overline{\text{MSE}}(\mu_z) \pm SD$	$\overline{\text{MSE}}(\sigma_z) \pm SD$	$\overline{\text{MSE}}(\mu_z) \pm SD$	$\overline{\text{MSE}}(\sigma_z) \pm SD$
0	0.028 ± 0.003	0.001 ± 0.000	0.119 ± 0.085	0.010 ± 0.009
0.42	0.032 ± 0.003	0.048 ± 0.002	0.065 ± 0.058	0.045 ± 0.005
0.85	0.542 ± 0.084	0.469 ± 0.022	0.549 ± 0.089	0.378 ± 0.004
1.70	1.369 ± 0.209	0.576 ± 0.006	1.396 ± 0.211	0.391 ± 0.009
3.40	2.878 ± 0.448	0.544 ± 0.014	2.932 ± 0.441	0.340 ± 0.010

Table 2: Posterior 95% credible interval of AUC between the tumor "core/annulus" and the healthy tissue for all subjects

Subject	"Core"			"Annulus"		
ID	90^{th}	95^{th}	97.5^{th}	90^{th}	95^{th}	97.5^{th}
1	(0.75,0.77)	(0.74,0.76)	(0.71,0.73)	(0.26,0.28)	(0.23,0.25)	(0.20,0.22)
2	(0.83,0.84)	(0.81,0.82)	(0.81,0.82)	(0.28,0.29)	(0.25,0.26)	(0.23,0.24)
3	(0.42,0.44)	(0.40,0.42)	(0.36,0.38)	(0.08,0.09)	(0.05,0.07)	(0.04,0.05)
4	(0.55,0.57)	(0.53,0.55)	(0.52,0.53)	(0.18,0.20)	(0.15,0.16)	(0.13,0.15)
5	(0.49,0.54)	(0.46,0.52)	(0.42,0.49)	(0.25,0.31)	(0.24,0.30)	(0.24,0.29)
6	(0.91,0.92)	(0.91,0.91)	(0.89,0.90)	(0.57,0.68)	(0.55,0.56)	(0.53,0.54)
7	(0.92,0.96)	(0.90,0.94)	(0.89,0.93)	(0.31,0.33)	(0.29,0.31)	(0.29,0.31)
8	(0.23,0.25)	(0.30,0.31)	(0.35,0.36)	(0.92,0.94)	(0.95,0.97)	(0.96,0.97)
9	(0.62,0.65)	(0.59,0.62)	(0.58,0.60)	(0.21,0.35)	(0.21,0.39)	(0.21,0.42)
10	(0.87,0.89)	(0.86,0.88)	(0.85,0.87)	(0.59,0.60)	(0.59,0.60)	(0.59,0.60)
11	(0.69,0.72)	(0.68,0.71)	(0.66,0.68)	(0.25,0.27)	(0.23,0.25)	(0.20,0.23)

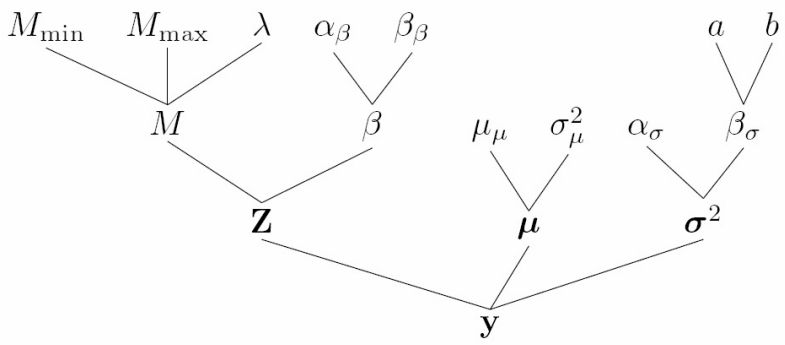

Figure 1: Directed Acyclic Graph (DAG) of the proposed model

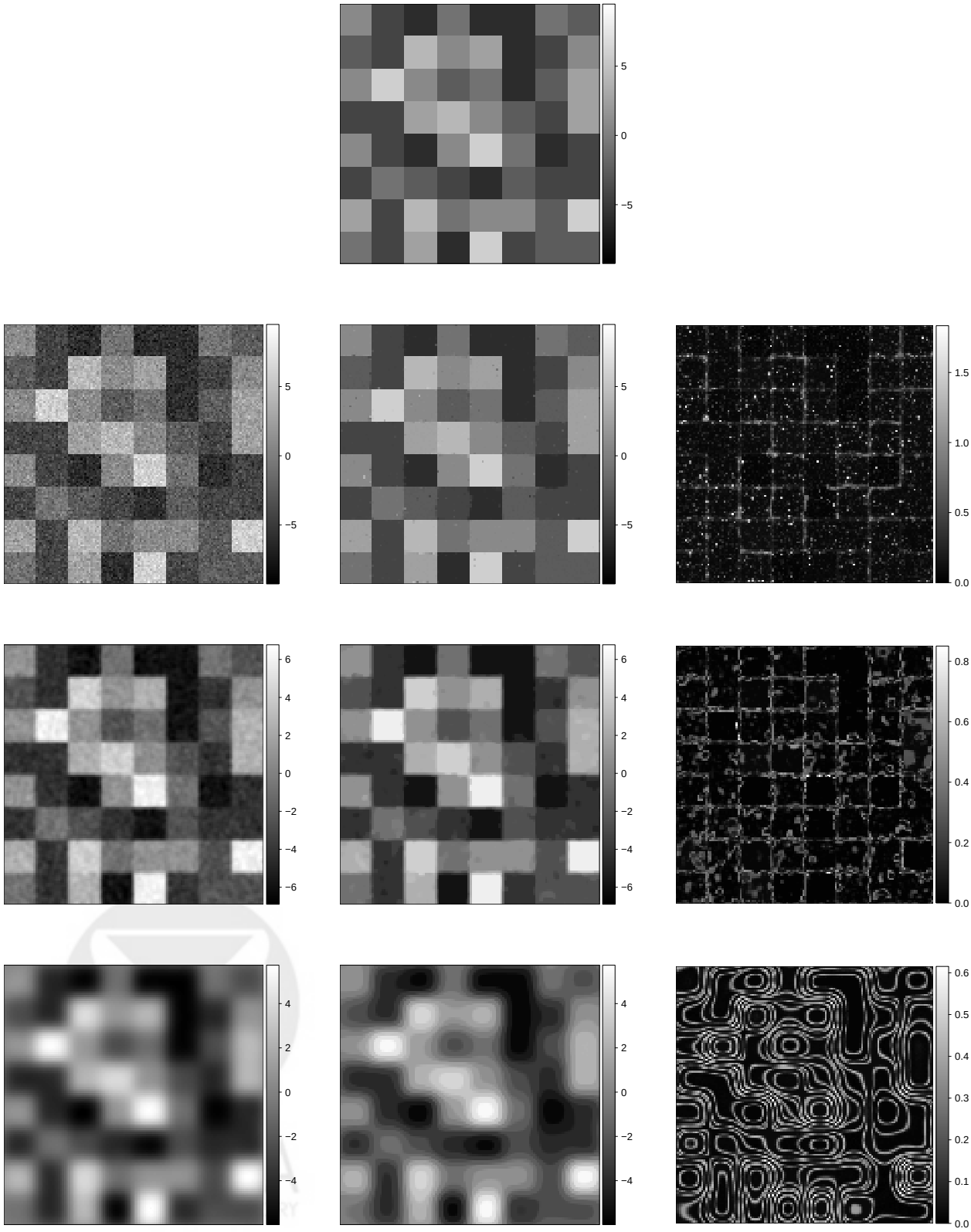

Figure 2: One set of simulations: the "true"scene ($\boldsymbol{\mu}_z^{\text{true}}$, top), the observed images ($\mathbf{y}$, left), the posterior mean ($\hat{\eta}_i$, middle) and standard deviation ($\hat{\psi}_i$, right) of pixel mean intensity when FWHM $= 0, 2$, and 8 from the second to fourth row.

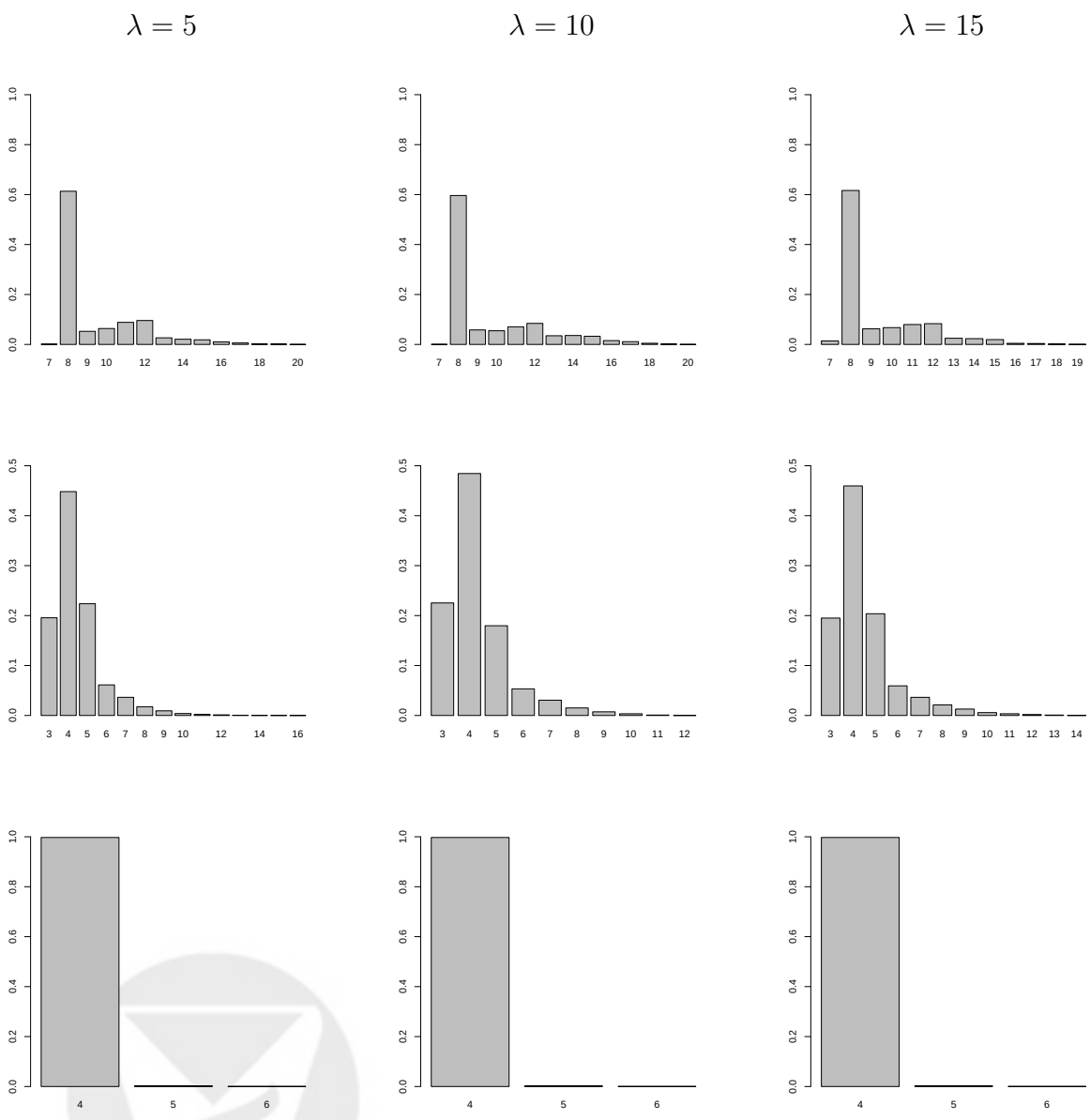

Figure 3: Sensitivity analysis on the hyperprior of number of components (λ), from top to bottom simulation study with conditionally independent noise, real data analysis for subject 1 and subject 2.

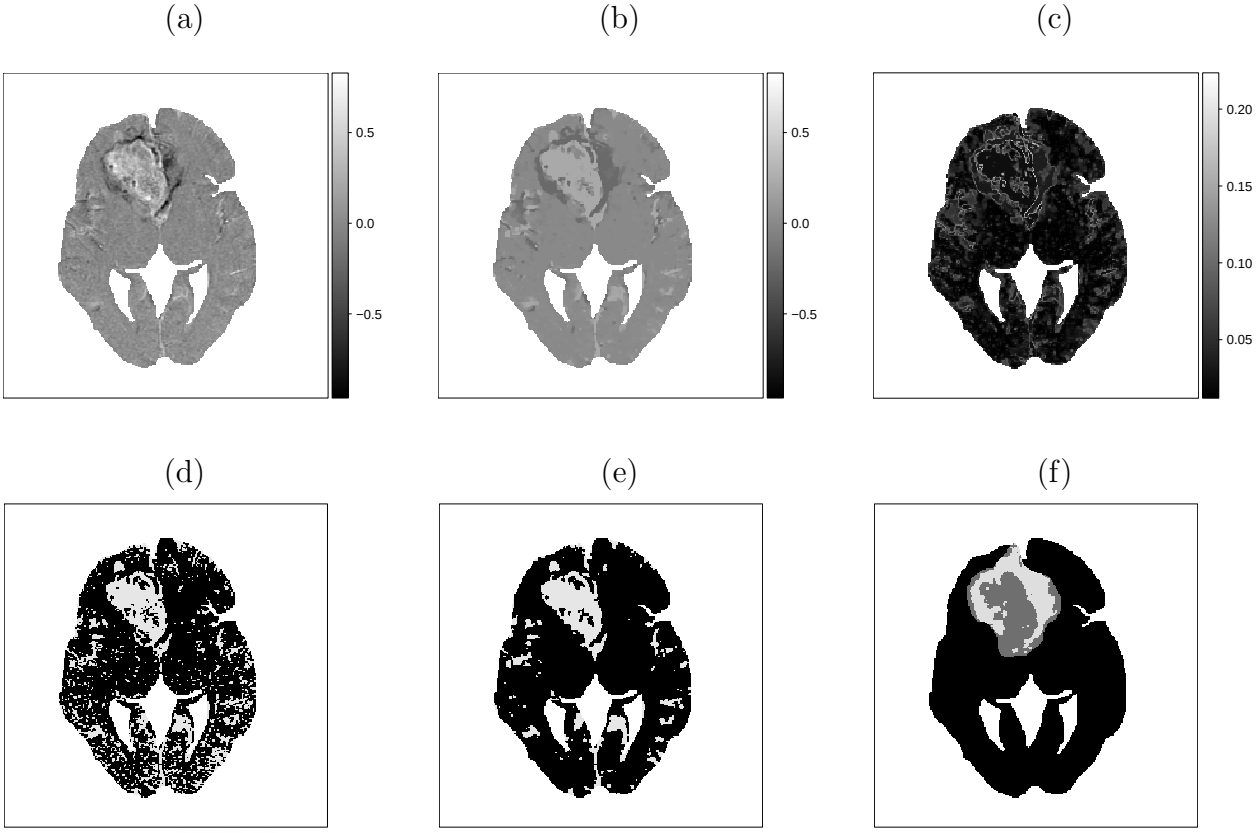

Figure 4: Results on subject 1: (a) observed change in contrast uptake (light shades standing for large increase); (b) marginal posterior mean of change in contrast uptake $\tilde{\mu}_{z_i}$ and (c) standard deviation $\tilde{\sigma}_{z_i}$; thresholded image (d) without and (e) with consideration of spatial structure; (f) tumor "core/annulus".

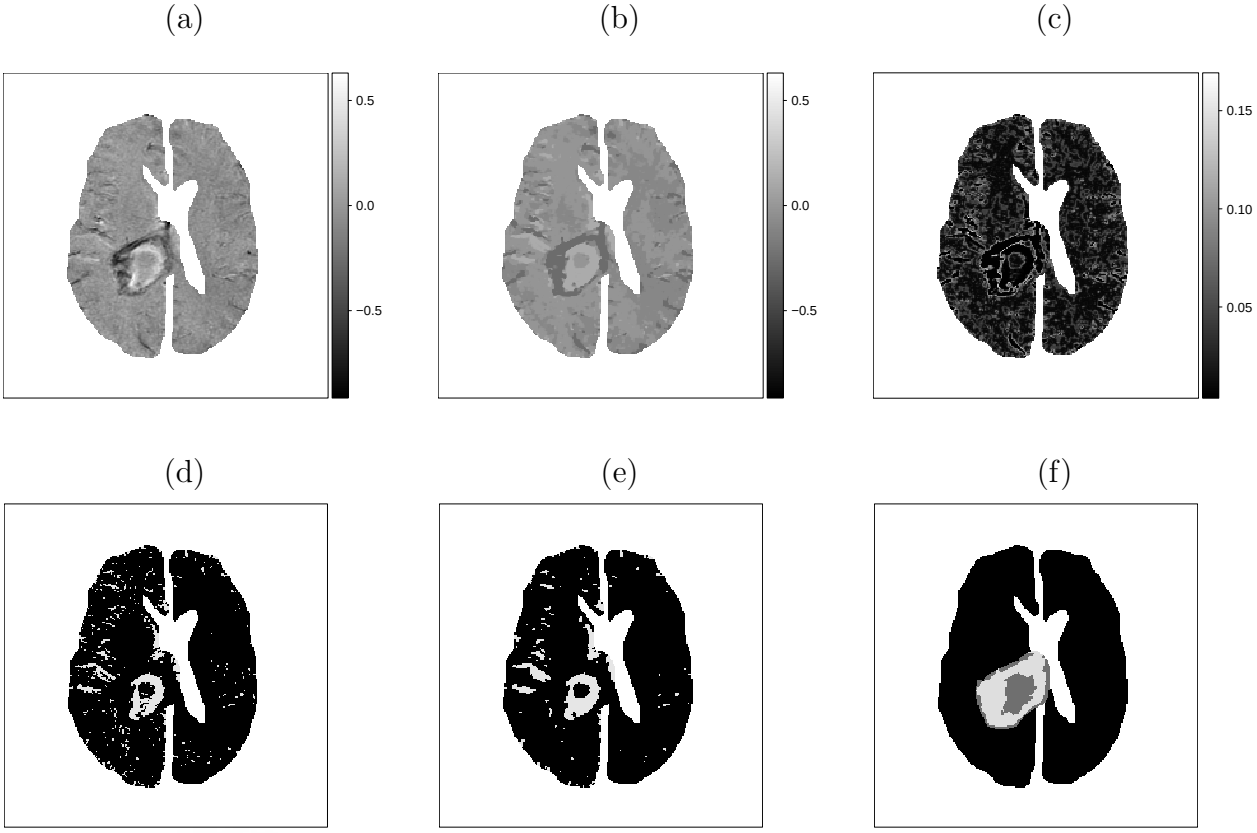

Figure 5: Results on subject 2: (a) observed change in contrast uptake (light shades standing for large increase); (b) marginal posterior mean of change in contrast uptake $\tilde{\mu}_{z_i}$ and (c) standard deviation $\tilde{\sigma}_{z_i}$; thresholded image (d) without and (e) with consideration of spatial structure; (f) tumor "core/annulus".