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

Year 2007 *Paper* 72

Quantitative Magnetic Resonance Image Analysis via the EM Algorithm with Stochastic Variation

Xiaoxi Zhang* Timothy D. Johnson†
Roderick J.A. Little‡

*

†tdjtdj@umich.edu

‡

Quantitative Magnetic Resonance Image Analysis via the EM Algorithm with Stochastic Variation

Xiaoxi Zhang, Timothy D. Johnson, and Roderick J.A. Little

Abstract

Quantitative Magnetic Resonance Imaging (qMRI) provides researchers insight into pathological and physiological alterations of living tissue, with the help of which, researchers hope to predict (local) therapeutic efficacy early and determine optimal treatment schedule. However, the analysis of qMRI has been limited to ad-hoc heuristic methods. Our research provides a powerful statistical framework for image analysis and sheds light on future localized adaptive treatment regimes tailored to the individual's response. We assume in an imperfect world we only observe a blurred and noisy version of the underlying "true" scene via qMRI, due to measurement errors or unpredictable influences. We use a hidden Markov Random Field to model the unobserved "true" scene and develop a maximum likelihood approach via the Expectation-Maximization algorithm with stochastic variation. An important improvement over previous work is the assessment of variability in parameter estimation, which is the valid basis for statistical inference. Moreover, we focus on recovering the "true" scene rather than segmenting the image. Our research has shown that the approach is powerful in both simulation studies and on a real dataset, while quite robust in the presence of some model assumption violations.

Quantitative Magnetic Resonance Image Analysis via the EM Algorithm with Stochastic Variation

Xiaoxi Zhang*, Timothy D. Johnson,

and

Roderick J. A. Little

Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109

**email:* xiaoxi@umich.edu

SUMMARY: Quantitative Magnetic Resonance Imaging (qMRI) provides researchers insight into pathological and physiological alterations of living tissue, with the help of which, researchers hope to predict (local) therapeutic efficacy early and determine optimal treatment schedule. However, the analysis of qMRI has been limited to ad-hoc heuristic methods. Our research provides a powerful statistical framework for image analysis and sheds light on future localized adaptive treatment regimes tailored to the individual's response. We assume in an imperfect world we only observe a blurred and noisy version of the underlying "true" scene via qMRI, due to measurement errors or unpredictable influences. We use a hidden Markov Random Field to model the unobserved "true" scene and develop a maximum likelihood approach via the Expectation-Maximization algorithm with stochastic variation. An important improvement over previous work is the assessment of variability in parameter estimation, which is the valid basis for statistical inference. Moreover, we focus on recovering the "true" scene rather than segmenting the image. Our research has shown that the approach is powerful in both simulation studies and on a real dataset, while quite robust in the presence of some model assumption violations.

KEY WORDS: EM algorithm; Hidden Markov Random Field; Missing Data; Model Selection; Quantitative MRI; Stochastic Variation.

1. Introduction

Quantitative Magnetic Resonance Imaging (qMRI) is a noninvasive tool for visualizing the inside of living organisms, and is used to assess pathological and physiological alterations in living tissue, such as the brain. More recently, it has been used to measure physiological changes (such as diffusion, perfusion, vascular permeability, and metabolism) in diseased tissue due to therapy (e.g. Cao et al. 2005; Moffat et al. 2005; Hamstra et al. 2005). With the aid of qMRI, investigators hope to predict (local) therapeutic efficacy early during treatment so that treatments can be tailored to the individual.

This work is motivated by a pilot qMRI study conducted at the University of Michigan School of Medicine. Eleven volunteers with primary, high-grade gliomas were recruited for the study. Prior to the initiation of radiation therapy, the volunteers underwent T1-weighted qMRI with and without contrast enhancement. The same imaging protocol was subsequently performed after approximately the first and third week of radiotherapy, and 1, 3, and 6 months after completion of radiotherapy. The contrast agent used was Gadolinium diethylenetriaminepentaacetic acid (Gd-DTPA). Gd-DTPA has a molecular diameter in the range of 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). This was the first study to use quantitative and high-resolution MRI to assess the effects of radiation on the vascular permeability in tumor and healthy tissue to a molecule in the size range of chemotherapeutic agents in high-grade gliomas (Cao et al. 2005).

Prognosis for primary high-grade gliomas is poor and advances in radiotherapy followed by chemotherapy have failed to prolong the median survival time of these patients beyond about 1 year after diagnosis. Chemotherapy has been largely unsuccessful due to the tight endothelial junctions in the tumor (blood-tumor barrier, BTB) that limit the delivery of these large chemotherapeutic molecules to the tumor cells. One of the goals of this study was to

determine the effects, over time, of radiation therapy on the 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 chemotherapeutic drugs, this may suggest an optimal time for delivering these drugs during radiation therapy, as opposed to waiting for the completion of radiation therapy. In this manuscript, we focus on the change in contrast uptake from the baseline imaging study to the 3-week imaging study, as this time point is of special interest to the investigators.

There is a large body of research on medical image analysis, in particular functional MRI (fMRI). The fMRI analyses are typically of a time series nature (Worsley and Friston 1995). Random Field Theory (RFT) is applied after smoothing the data with an isotropic Gaussian filter. It assumes spatial continuity, which is reasonable for healthy volunteers in fMRI experiments. However, qMRI in the motivating study was performed on patients with solid mass tumors, where distinct spatial discontinuities between healthy tissue and the tumor are visible (Figures 2a and 3a). The tumors are physiologically different from surrounding healthy tissue, and their contrast uptake is highly heterogeneous. The smoothing techniques that are applied in most image analyses blur tissue boundaries, and hence do not model this feature well. Furthermore, 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 can lead to incorrect variance estimates and invalid hypothesis tests.

In this manuscript we analyze qMRI data using a flexible, yet conceptually simple hidden Markov Random Field model (MRF, Besag 1974), also known as the Potts model in the statistical physics literature (Potts 1952). Our model avoids over-smoothing of the data while accounting for spatial correlation. The rationale for this model is that qMRI produces a blurred and noisy version of the underlying change in contrast uptake, due to measurement

error, reconstruction error, field inhomogeneities and other unpredictable influences. The underlying image, which we call the "true" scene, is represented as a finite number of homogeneous regions. We assign discrete labels to the regions with the same change in contrast uptake, and use a hidden MRF to model the spatial layout of the unobserved labels. To fit the model, we treat these unobserved MRF labels as missing data, and compute maximum likelihood (ML) estimates using the Expectation Maximization (EM, Dempster, Laird, and Rubin 1977) algorithm with stochastic variation (Wei and Tanner 1990).

Previous implementations of the MRF model have problems or limitations. Either the likelihood was incorrectly specified, the EM algorithm was incorrectly implemented, or the spatial regularization parameter in the MRF was assumed to be known; results are highly sensitive to the choice of this parameter. Some researchers (Chalmond 1989; Won and Derin 1992; Zhang, Modestino, and Lagan 1994; and Panjwani and Healey 1995) used the pseudo-likelihood approach proposed by Besag (1974). Melas and Wilson (2002) pointed out that the pseudo-likelihood approach tends to overestimate the regularization parameter of the MRF and over-smooth the data. Zhang, Brady, and Smith (2001) and Sengur, Turkoglu, and Ince (2006) correctly specified the likelihood, but incorrectly implemented the EM algorithm as we discussed later. Lei and Udupa (2003) used the Iterative Conditional Mode (ICM, Besag 1974, Besag 1986) algorithm, which they refer to as "MRF-ICM". However, they incorrectly wrote the joint distribution of the MRF as the product of local conditional distributions, which in a sense emulates Besag's pseudo-likelihood approach. These approaches maximizing the complete-data likelihood jointly with respect to parameters and missing data, an approach which in general lacks the consistency and asymptotic efficiency of ML (Little and Rubin 1983). Furthermore, estimates of uncertainty of the parameter estimates and the hidden MRF labels are not provided (see also Deng and Clausi 2004). The goal of most image analyses via MRFs is segmentation, and therefore aims at a labeling of all pixels. This optimization task

is often performed in an iterative fashion (Won and Derin 1992), or via simulated annealing (Lakshmanan and Derin 1989). Segmentation is not of primary interest in our application. In fact, the segmentation labels lack a strong biologically meaningful interpretation. Rather, we are using the hidden MRF model as a way to account for spatial correlation and as a way to perform edge-preserving smoothing of the image. It is the unobserved "true" changes that are of scientific interest.

In this manuscript, we build on previous work on the hidden MRF model by 1) correctly implement the EM algorithm with stochastic variation; 2) estimating the spatial regularization parameter rather than assuming that it is known; and 3) estimating standard errors of the parameter estimates via the Louis (1982) method. We focus on estimating the local mean change in contrast uptake (the "true" scene) rather than segmentation.

This manuscript is organized as follows. In Section 2, we present our model and the EM algorithm, and discuss estimated standard errors and model selection. In Section 3, we present results from a simulation study where we investigate the sensitivity to model assumption violations. Results from the motivating example are presented in Section 4. We conclude by summarizing the strengths and weaknesses of our approach, and discussing future work.

2. Model and Algorithm

2.1 *Image Model*

We use the following notation. Pixels (short for picture elements) will be indexed by $i = 1, 2, \cdots, N$. If pixel i and i' are immediately adjacent (sharing a common edge), we call them neighbors, denoted $i \sim i'$. The set of neighbors of pixel i is denoted $\partial i = \{i' : i' \sim i\}$. Associated with each pixel i are the observed pixel intensity y_i and a hidden label z_i. The collection of the observed pixel intensities $\mathbf{y}^{\mathrm{T}} = (y_1, y_2 \cdots, y_N)$ is called the image (i.e.

the change in contrast uptake), while the collection of latent labels $\mathbf{z}^{\mathrm{T}} = (Z_1 = z_1, Z_2 = z_2, \cdots, Z_N = z_N)$ defined on a finite discrete state space 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.

We assume there is an M-state MRF on the state space $\mathcal{S} = \{1, 2, \cdots, M\}$. Each state is mapped to an intensity in the "true" scene. It follows that there are M^N configurations on the configuration space $\mathcal{S}^N$, the number of which increases exponentially with the number of pixels N. Our image model is a two-level hierarchical model. The higher level specifies the spatial structure of the MRF with probability mass function

$$\Pr(\mathbf{Z} = \mathbf{z} \mid \beta) = g^{-1}(\beta) \exp\Big\{ \sum_{i \sim i'} \beta \, \mathrm{I}(z_i = z_{i'}) \Big\}, \text{ for all } \mathbf{z} \in \mathcal{S}^N,$$

where $\mathrm{I}(\cdot)$ is the indicator function and the regularization parameter $\beta \geqslant 0$ controls the spatial smoothness of the MRF. The normalizing constant

$$g(\beta) = \sum_{\mathbf{z} \in \mathcal{S}^{\mathbf{N}}} \exp\Big\{ \sum_{i \sim i'} \beta \, \mathrm{I}(z_i = z_{i'}) \Big\}$$

has M^N summands, and is not analytically tractable. Given $\mathbf{Z} = \mathbf{z}$, the observed pixel intensities are conditionally independent with Gaussian noise on the lower level,

$$y_i \mid z_i = \mu_{z_i} + e_i, \quad e_i \sim \mathrm{N}(0, \sigma^2_{z_i}), \text{ for all } i.$$

We write $\boldsymbol{\mu} = (\mu_1, \cdots, \mu_M)^{\mathrm{T}}$ and $\boldsymbol{\sigma}^2 = (\sigma^2_1, \cdots, \sigma^2_M)^{\mathrm{T}}$, and $\boldsymbol{\theta}^{\mathrm{T}} = (\boldsymbol{\mu}^{\mathrm{T}}, \boldsymbol{\sigma}^{2\mathrm{T}}, \beta)$.

A few comments are in order: 1) The model requires little prior knowledge about the spatial structure of the hidden configuration, only that neighboring pixels tend to share the same label. 2) The regularization parameter, β, controls the strength of the association between neighbors. 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. Note that when $\beta = 0$ our model degenerates to a non-spatial Gaussian mixture model with equal component weights, in which case pixels are independent. The spatial correlation decreases

as the distance between pixels increases. 3) Although the likelihood assumes conditionally independent Gaussian noise given the hidden labels, the data are marginally dependent.

2.2 *The EM Algorithm with stochastic variation*

With the introduction of the unobservable labels in the hidden MRF, the image model can be viewed as a missing data problem. The changes in contrast uptake are the observed data ($Y_{\text{obs}} = \mathbf{y}$) and the pixel labels are treated as missing data ($Y_{\text{mis}} = \mathbf{Z}$). This is in contrast to most existing frequentist analyses that treat both the model parameters and the hidden labels as unknown but fixed quantities. The complete-data log-likelihood is

$$\begin{aligned} l_{\text{comp}}(\boldsymbol{\theta}) &= \log f(\mathbf{y} \mid \mathbf{Z} = \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\sigma}^2) + \log \Pr(\mathbf{Z} = \mathbf{z} \mid \beta) \\ &= -0.5N \log(2\pi) - \sum_{k=1}^{M} \sum_{i \in D_k} \left\{ \log(\sigma_k) + 0.5\sigma_k^{-2}(y_i - \mu_k)^2 \right\} \\ &\quad + \sum_{i \sim i'} \beta \mathrm{I}(z_i = z_{i'}) - \log g(\beta). \end{aligned}$$

This belongs to the exponential family with complete data sufficient statistics $T_{k1} = N_k$, $T_{k2} = \sum_{i \in D_k} y_i$, $T_{k3} = \sum_{i \in D_k} y_i^2$ for $k = 1, 2, \cdots, M$, and $T_4 = \sum_{i \sim i'} \mathrm{I}(z_i = z_{i'})$. The component with common label k is denoted as $D_k = \{i : z_i = k\}$ with N_k pixels.

We maximize the log-likelihood via the EM algorithm. The EM algorithm iterates between an E-step (expectation) and an M-step (maximization). At the t'th iteration, the E-step computes the conditional expectation of the complete-data sufficient statistics given the observed data and current parameter estimates $\theta^{(t)}$. The M-step updates $\theta = \theta^{(t+1)}$ as the solution of the complete-data likelihood equations. An advantage of EM is that it avoids the computation of the normalizing constant $g(\beta)$.

The conditional expectation in the E-step in our setting is

$$\mathbf{T}^{(t)} = \sum_{\mathbf{z} \in \mathcal{S}^{\mathbf{N}}} \mathbf{T}(\mathbf{y}, \mathbf{Z} = \mathbf{z}) \Pr\left(\mathbf{Z} = \mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}^{(t)}\right), \tag{1}$$

where $\mathbf{T}(\mathbf{y}, \mathbf{Z} = \mathbf{z}) = (T_{11}, T_{12}, T_{13}, \cdots, T_{M1}, T_{M2}, T_{M3}, T_4)$, and $\Pr\left(\mathbf{Z} = \mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}^{(t)}\right)$ is the conditional distribution of the latent labels. It has M^N summands, and is not analytically

tractable. One solution is to stochastically approximate the expectations in the E-step as in Monte Carlo EM (MCEM, Wei and Tanner 1990). The t'th EM iteration consists of the following steps:

(1) E-step: Draw configurations $\mathbf{z}^{(1)}, \cdots, \mathbf{z}^{(S_t)} \sim \Pr\left(\mathbf{Z} \mid \mathbf{y}, \boldsymbol{\theta}^{(t)}\right)$, and compute the Monte Carlo estimates of the conditional expectation of the sufficient statistics $\mathbf{T}(\mathbf{y}, \mathbf{z})$ given the observed data $\mathbf{y}$ and current parameter estimates $\boldsymbol{\theta}^{(t)}$,

$$T_{k1}^{(t)} = S_t^{-1} \sum_{s=1}^{S_t} N_k^{(s)},\ T_{k2}^{(t)} = S_t^{-1} \sum_{s=1}^{S_t} \Big(\sum_{i \in D_k^{(s)}} y_i \Big),\ T_{k3}^{(t)} = S_t^{-1} \sum_{s=1}^{S_t} \Big(\sum_{i \in D_k^{(s)}} y_i^2 \Big),$$

$$\text{for } k = 1, \cdots, M, \text{ and } T_4^{(t)} = S_t^{-1} \sum_{s=1}^{S_t} \Big\{ \sum_{i \sim i'} \mathrm{I}(z_i^{(s)} = z_{i'}^{(s)}) \Big\}.$$

We draw $\mathbf{z}^{(s)}$ using the Swendsen-Wang algorithm (Swendsen and Wang, 1987), an efficient sampler specifically developed for the Potts model. It updates labels for clusters of pixels rather than one pixel at a time as in ICM.

Since the complete-data log-likelihood consists of two distinct parts $l(\boldsymbol{\mu}, \boldsymbol{\sigma}^2)$ and $l(\beta)$, the M-step has two parts:

(2) M1-step: update the Gaussian parameter estimates $\left(\boldsymbol{\mu}^{(t+1)}, \boldsymbol{\sigma}^{2(t+1)}\right)$ based on the expections in the E-step,

$$\mu_k^{(t+1)} = T_{k2}^{(t)} / T_{k1}^{(t)},\ \sigma_k^{2(t+1)} = T_{k3}^{(t)} / T_{k1}^{(t)} - \left(\mu_k^{(t+1)}\right)^2,\ \text{for } k = 1, 2, \cdots, M.$$

(3) M2-step: solve the regularization parameter estimate $\beta^{(t+1)}$ from

$$\frac{\partial l_{\text{comp}}}{\partial \beta} = 0,\ \text{i.e. } g'(\beta) / g(\beta) = T_4^{(t)}.$$

Since the ratio $g'(\beta)/g(\beta)$ is a monotone function of β (proof see the Appendix), any root finding algorithm can be used for this step.

2.3 *Standard Errors of Parameter Estimates*

We use the method of Louis (1982) to derive the asymptotic covariance matrix of the parameter estimates based on the observed loglikelihood $l_{\text{obs}} = \log f(\mathbf{y} \mid \boldsymbol{\theta})$

$$\mathrm{I}_{\text{obs}}^{-1} = -\left(E_{\boldsymbol{\theta}} \frac{\partial^2 l_{\text{obs}}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right)^{-1}.$$

The observed information matrix is formulated as the difference of the complete-data information I_{comp} and the information for the conditional distribution of missing data given the observed data I_{mis}, i.e.

$$\begin{aligned} \mathrm{I}_{\text{obs}}(\boldsymbol{\theta}) &= \mathrm{I}_{\text{comp}}(\boldsymbol{\theta}) - \mathrm{I}_{\text{mis}}(\boldsymbol{\theta}) \\ &= -E_{\boldsymbol{\theta}} \left(\frac{\partial^2 l_{\text{comp}}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \Big| \mathbf{y} \right) - E_{\boldsymbol{\theta}} \left(\frac{\partial l_{\text{comp}}}{\partial \boldsymbol{\theta}} \frac{\partial l_{\text{comp}}}{\partial \boldsymbol{\theta}^T} \Big| \mathbf{y} \right) \\ &\quad + E_{\boldsymbol{\theta}} \left(\frac{\partial l_{\text{comp}}}{\partial \boldsymbol{\theta}} \Big| \mathbf{y} \right) E_{\boldsymbol{\theta}} \left(\frac{\partial l_{\text{comp}}}{\partial \boldsymbol{\theta}^T} \Big| \mathbf{y} \right). \end{aligned}$$

It follows from the asymptotic properties of the MLE that $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \sim \mathrm{N}\big(0, \mathrm{I}_{\text{obs}}^{-1}(\hat{\boldsymbol{\theta}})\big)$.

2.4 *Estimating the Number of States in the Markov Random Field*

Since there is no clear substantive rationale for determining the number of states in the MRF, we use information criteria, such as the Akaike Information Criterion (AIC, Akaike 1973) and the Bayesian Information Criterion (BIC, Schwarz 1978). We run the proposed algorithm for a range of values of M, and compute

$$\begin{aligned} \mathrm{AIC}_M &= -2 l_{\text{obs}}(\hat{\boldsymbol{\theta}}_M \mid \mathbf{y}, M) + 2(2M+1) \\ \mathrm{BIC}_M &= -2 l_{\text{obs}}(\hat{\boldsymbol{\theta}}_M \mid \mathbf{y}, M) + (2M+1) \log N, \end{aligned}$$

where $l_{\text{obs}}(\hat{\boldsymbol{\theta}}_M \mid \mathbf{y}, M)$ is the observed log-likelihood of the model with an M-state MRF. Smaller AIC or BIC is preferred – there does not appear to be a consensus choice between these criteria, but in our application they lead to the same value of M. We use proper multiple imputation (Rubin 1987) to approximate $l_{\text{obs}}(\hat{\boldsymbol{\theta}}_M \mid \mathbf{y}, M)$, via the expression

$$\hat{l}_{\text{obs}}(\hat{\boldsymbol{\theta}}_M \mid \mathbf{y}, M) \approx D^{-1} \sum_{d=1}^{D} l_{\text{comp}}(\hat{\boldsymbol{\theta}}_M \mid \mathbf{y}, \mathbf{z}_M^{(d)}, M),$$

where $\boldsymbol{\theta}_M^{(d)} (d = 1, 2, \cdots, D)$ are drawn from the asymptotic distribution $\mathrm{N}\big(\hat{\boldsymbol{\theta}}_M, \mathrm{I}_{\mathrm{obs}}^{-1}(\hat{\boldsymbol{\theta}}_M)\big)$, and $\mathbf{z}^{(d)}$ are drawn from $\Pr(\mathbf{Z}_M \mid \boldsymbol{\theta}_M^{(d)}, \mathbf{y}, M)$.

3. Results

We first conduct a simulation study under the model assumptions to evaluate the performance of the proposed method. We then apply the algorithm when the observed image is smoothed with various Gaussian smoothing kernels, to assess robustness to violations of the model assumption of conditional independence. The variance of the Gaussian kernels are specified as the spread of an (unnormalized) density at half of its maximum value (the full width at half maximum, FWHM), an approach commonly employed in signal processing. The FWHM is related to the standard deviation of a Gaussian distribution via the expression $\sigma = \mathrm{FWHM}/(2\sqrt{2\log 2})$. A large FWHM corresponds to wide bandwidth and results in heavy smoothing and large spatial correlation. In the simulation studies, we use a superscript to denote the FWHM of the Gaussian kernel used (i.e. $\mathbf{y}^{\mathrm{FWHM}}$), and $\mathbf{y}^0$ means no smoothing (i.e. conditionally independent noise). The results on the real dataset are also presented.

3.1 *Simulation Study with Conditionally Independent Noise*

We simulate a hidden configuration ($\mathbf{z}^{\mathrm{true}}$) with ten distinct components on a 128×128 lattice, and map it into a "true" scene $\boldsymbol{\mu}_{\mathbf{z}^{\mathrm{true}}} = (\mu_{z_1^{\mathrm{true}}}, \mu_{z_2^{true}}, \cdots, \mu_{z_N^{\mathrm{true}}})$ (Figure 1 top panel). We use light gray to denote high intensity. Under the conditionally independent noise assumption, we add white noise to $\boldsymbol{\mu}_{\mathbf{z}^{\mathrm{true}}}$ to obtain the observed image, $\mathbf{y}^0$. The parameters used to generate the simulation are listed in Table 1. The signal-to-noise ratio (SNR), defined as

$$\mathrm{SNR} = \frac{\text{difference in component mean}}{\text{standard deviation of the noise}},$$

is 1.70.

We first fit a Gaussian mixture model (ignoring the spatial structure) with 10 components and equal component weights. Over half of the pixels are misclassified due to the high noise

level. There is also considerable bias in the parameter estimates (Table 1). Next, we fit our proposed algorithm with $M = 6, 7, \cdots, 16$, components. The initial values of μ_k, $k = 1, 2, \cdots, M$, are evenly spaced over the range of the data $(-11.2, 12.6)$. The initial values of σ_k, $k = 1, 2, \cdots, M$, are $1/(2M)$ times the range of the data. Ten components are selected by both AIC and BIC. The final estimates $\hat{\boldsymbol{\theta}}_{M=10}$ and their standard deviations are listed in Table 1. All 95% confidence intervals cover the true parameter values. We estimated the vector of pixel intensities, $\mu_{z_i}^{\text{est.}} = E(\hat{\mu}_{Z_i} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_{M=10})$, using Monte Carlo estimates based on 500 samples of the hidden configuration drawn from $\Pr(\mathbf{Z} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_{M=10})$ (Figure 1 second row). As a measure of the difference between μ_z^{true} and the estimated pixel intensity $\widehat{\mu}_{z_i}^{\text{est.}}$, we compute the sum of squared discrepancies between them, $SS_{<\text{est,true}>} = \sum_{i=1}^{N} \left(\hat{\mu}_{z_i^{est.}} - \mu_{z_i^{true}} \right)^2$ (smaller value suggests better fit). The sum of squared discrepancy of the proposed algorithm is 298.33, less than 2% of the sum of squares of the noise $SS_{<\text{obs}^0,\text{true}>} = \sum_{i=1}^{N} (y_i^0 - \mu_{z_i}^{\text{true}})^2 = 16539.63$.

We also investigate the false positive rate (FPR) and false negative rate (FNR) under the simulation study. The proposed algorithm uniformly produces lower FPR and FNR than ignoring spatial correlation. Choosing an arbitrary threshold value of 5.0, the FPR and FNR of directly thresholding the observed image $\mathbf{y}^0$ are 3.1% and 9.8%, compared to 0.1% and 0% when considering the spatial structure.

[Table 1 about here.]

[Figure 1 about here.]

Zhang et at. (2001) implemented an EM-type algorithm for a similar model to ours, and we implement it for comparison. In their E-step, they compute the function

$$\begin{aligned}
\tilde{Q}\left(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\right) &= \sum_{i=1}^{N} \sum_{k=1}^{M} \Pr\left(Z_i = k \mid y_i, \boldsymbol{\theta}^{(t)}\right) \log f\left(y_i, Z_i = k \mid \mathbf{z}_{\partial i}, \boldsymbol{\theta}^{(t)}\right) \\
&= \sum_{i=1}^{N} \sum_{k=1}^{M} \Pr\left(Z_i = k \mid y_i, \boldsymbol{\theta}^{(t)}\right) \left\{ \log f\left(y_i \mid Z_i = k, \theta^{(t)}\right) + \log \Pr\left(Z_i = k \mid \mathbf{z}_{\partial i}, \boldsymbol{\theta}^{(t)}\right) \right\}.
\end{aligned}$$

A correct implementation of the E-step computes the expectation of the conditional log-like

lihood given the observed data, namely

$$\begin{aligned} Q\big(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\big) &= \sum_{\mathbf{z}\in\mathcal{S}^{\mathbf{N}}} \Pr\big(\mathbf{Z}=\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}^{(t)}\big) \log f\big(\mathbf{y}, \mathbf{z} \mid \boldsymbol{\theta}^{(t)}\big) \\ &= \sum_{\mathbf{z}\in\mathcal{S}^{\mathbf{N}}} \Pr\big(\mathbf{Z}=\mathbf{z} \mid \mathbf{y}, \boldsymbol{\theta}^{(t)}\big) \Big\{\log f\big(\mathbf{y} \mid \mathbf{z}, \boldsymbol{\theta}^{(t)}\big) + \log \Pr\big(\mathbf{Z}=\mathbf{z} \mid \boldsymbol{\theta}^{(t)}\big)\Big\}. \end{aligned}$$

Note that $Q\big(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\big)$ differs from $\tilde{Q}\big(\boldsymbol{\theta} \mid \boldsymbol{\theta}^{(t)}\big)$ in that the summation in Q extends over $M \times N$ terms.

The misclassification rate (MCR) from the last configuration from their algorithm is 3.7%, compared with 1.6% for our algorithm. The estimated configuration from their approach has many more small patches of "incorrect" labels. Moreover, their parameter estimates generally have larger bias than our proposed method, especially in the standard deviation estimates (Table 1). We also notice that the configuration in their approach gets stuck after a few dozen iterations, which we think indicates slow mixing and a tendency to get trapped in local modes. As a matter of fact, when we initialize our algorithm with their final configuration and parameter estimates, the loglikelihood always increases. For instance in one run it increases from -40811 to -40533.

3.2 *Simulation Studies with Correlated Noise*

After some algebraic manipulations, one can see that, under the conditionally independent noise assumption, the observed intensities are marginally dependent (i.e. $\mathrm{Corr}(y_i, y_{i'}) \neq 0$). This is due to the spatial correlation induced by the MRF. However, conditional independence is still a strong assumption. Therefore, we conduct a series of simulation studies with varying degrees of correlated noise to assess the robustness of our model to violations of this model assumption.

We apply Gaussian smoothing kernels with FWHM $= 1, 2, 4$, and 8 ($\sigma = 0.42, 0.85, 1.70$, and 3.40) on $\mathbf{y}^0$. Due to space limitations, we display the results using FWHM $= 2$ and 8 while fixing $M = 10$ (Figure 1 bottom two rows). The "edge preservation" of the proposed

method is evident, especially when FWHM = 8 (bottom row in Figure 1). However, some local features are not recovered due to extensive smoothing (e.g. the corners are rounded). The images in the third column in Figure 1 display the standard deviation of the estimated pixel intensities. The larger errors occur near the boundaries of classes. Although some smoothing is intrinsic in qMRI reconstruction algorithms, the above results suggest that the common practice of smoothing before image analysis for noise-reduction purposes is not necessarily when combined with our proposed method. Quantitatively, when the smoothing is relatively local (FWHM = 1), the estimated intensities are only slightly worse than with no smoothing ($SS_{<\text{est}^1,\text{true}>} = 251.87$). When the smoothing is more global (FWHM $= 2, 4, 8$), the sum of squared discrepancy are large, $SS_{<\text{est}^2,\text{true}>} = 2230.77, SS_{<\text{est}^4,\text{true}>} = 5592.80$, and $SS_{<\text{est}^8,\text{true}>} = 12140.11$.

3.3 *Application*

In the motivating study, eleven patients 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 use a subset of the data, i.e. the pre-radiation visit and the visit at approximately 3 weeks after the initiation. We take the change in contrast uptake from the baseline to the 3-week follow-up visit as a surrogate of the change in vascular permeability (Figure 2a and 3a), which is of special interest to the investigator. To save space, we only display the results on two patients. The other patients demonstrate similar results.

We first ignore the spatial information and pool all pixel intensities within each subject. The observed change in contrast uptake in the tumor has much heavier tails than in the

healthy tissue. Although a two sample t-test suggests a statistically significant difference ($p < 0.0001$) between the tumor (mean 0.017 for subject 1 and -0.080 for subject 2) and the healthy tissue (mean 0.008 for subject 1 and -0.053 for subject 2) for both patients, the absolute difference in means is quite small—significance is driven by the extremely large number of pixels and is most likely uninteresting. More importantly, it does not provide information on the differential change in contrast uptake between the tumor and healthy tissue.

We ran the proposed algorithm using several different numbers of hidden states: $M = 2, 3, \cdots, 14$. Both AIC and BIC choose $M = 3$ as the best model for subject 1. The three component mean estimates are -0.187 ± 0.004 (mean $\pm$ standard deviation), 0.005 ± 0.0005 and 0.210 ± 0.003 respectively. A negative value indicates a decrease in contrast uptake, while a positive value indicates an increase. As stated earlier, we are interested in the "true" changes in contrast uptake, rather than the hidden labels, which lack a biological interpretation. Therefore, we display the estimated pixel intensities $\mu_{z_i}^{est.} = E(\hat{\mu}_{Z_i} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_{M=3})$ in Figure 2b, which clearly shows the two concentric rings in the observed image (Figure 2a). Pixels near the boundaries of components are more variable than those away from the boundaries (Figure 2c).

[Figure 2 about here.]

The results from patient 2 are similar (Figure 3). Both AIC and BIC favor $M = 4$. The four component mean estimates are -0.264 ± 0.004, -0.095 ± 0.001, -0.013 ± 0.001 and 0.120 ± 0.004 respectively.

[Figure 3 about here.]

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. A biologically meaningful threshold of change has not been defined in this exploratory study, but for illustrative purpose, we choose a threshold of 0.06.

Ignoring the spatial structure of the data and thresholding the observed change in uptake (patient 1), regions that lie above the threshold scatter throughout the tumor/brain (41.2% of the tumor and 19.9% of the healthy tissue, Figure 2d), much of which, we believe, is attributable to random noise. Our proposed algorithm borrows strength from neighboring pixels, reducing both FPR and FNR and producing a smoother image (Figure 2e). Overall, 39.4% of the tumor exceeds the threshold while only 7.6% of healthy tissue exceeds the threshold. For patient 2, we found that 29.8% of the tumor exceeds the threshold as compared to 3.9% of healthy tissue. The thresholded image is again smoother when we account for spatial correlation (Figure 3e) than when we ignore it (27.3% of the tumor and 5.0% of the health tissue exceed the chosen threshold, Figure 3d).

In the original analysis, Cao et al. (2005) divided the tumor into two regions. One region where the pre-treatment contrast uptake was relatively large and the second region where it was relatively small. These regions typically divided the tumor into a "core" (low initial contrast uptake) surrounded by an annulus (high initial contrast uptake). The biological rationale for this is that the core of the tumor is typically hypoxic (low oxygen content) due to a lack of blood supply, while the annulus of the tumor is rich in blood supply due to angiogenesis (new blood vessel growth, that is typically disorganized in tumors and thus leaky). Hypoxia is known to have a protective effect against damage due to both radiation and chemotherapy. Thus, both radiation and chemotherapy will have some effect on the annulus

of the tumor and relatively little effect on the hypoxic core—which is then a source of tumor regrowth. Hence, the focus was on demonstrating that radiation therapy has a transient, albeit small, effect on the core of the tumor with respect to increasing the contrast uptake. If one could predict when this small increase takes place, it may then provide rationale for initiating chemotherapy; thus allowing for more effective control of the tumor core.

To divide the tumor into these two regions, Cao et al. (2005) use one standard deviation above the average contrast uptake in the healthy tissue that received a total dose less than 10 Gy. This number was then used to divide the tumor into initially enhanced (high contrast uptake) and initially non-enhanced regions (low contrast uptake). This criterion is arguably low. Furthermore, a large portion of healthy tissue, that receiving more than 10 Gy, is ignored, as is the spatial correlation inherent in the data. We take a different approach to dividing the tumor. First, we run the proposed algorithm on the baseline contrast uptake image and divide the tumor into initially enhanced and non-enhanced regions based on the 95^{th} percentile of the estimated healthy tissue contrast uptake. As seen in Figure 2f and 3f, the initially enhanced tumor area (in light gray shade) roughly corresponds to an annulus surrounding the non-enhanced area—the core (in medium gray shade). We also note that, in both patients, there is a thin outer annulus of non-enhancing tumor. We suspect that this may be caused by two sources of error. One, the tumor outlines were obtained from a radiation oncologist for radiation planning purpose and therefore may contain a thin margin outside the observed tumor region (to ensure that all the tumor received a uniform dose of radiation). Secondly, this may be caused by volume averaging (pixels near the edge of the tumor contain both diseased and healthy tissue). Nevertheless, we include this region as part of the initially non-enhanced region in our analysis.

For patient 1, 59.1% of the initially non-enhanced tumor region has an increase in contrast uptake above the threshold, 0.06, compared to 14.4% in the initially enhanced region (Figure

4). Similarly, for patient 2, 54.5% of the initially non-enhanced tumor region has a change in uptake that exceeds the threshold, compared to 12.6% in the enhanced region (Figure 5).

[Figure 4 about here.]

[Figure 5 about here.]

Admittedly, the choice of 95^{th} percentile for defining the initially enhanced and non-enhanced tumor regions is ad-hoc. We performed a small study to address the sensitivity of our results to this choice. Our tumor division was based on the 95^{th} percentile of the contrast uptake in healthy tissue. We compared our results to those using the 90^{th} and 97.5^{th} percentiles. The percentages of the initially enhanced and non-enhanced regions of the tumors that exceed the various thresholds are given in Table 2. From this table, it is evident that our results are not highly sensitive to the choice of threshold over the range of thresholds studied.

[Table 2 about here.]

4. Conclusion

We have proposed an image smoothing algorithm suitable for qMRI data with edge preservation. Compared to previous work on similar models, we show how to correctly implement the stochastic variation of the EM algorithm. More importantly, we quantify the uncertainty in parameter estimates; previous work targets the hidden labels and point estimation of parameters without attempting to quantify this uncertainty. Furthermore, we emphasize the estimate of the true change in contrast uptake rather than the hidden labels, which are hard to interpret. The performance of the proposed method is satisfactory in both simulation studies and on real data. The model is rather robust to violations of the model assumption. Although the algorithm works fine under moderate smoothing of the observed image as

shown in Section 3.2, the results degrade as the smoothing becomes heavy. We therefore suggest that no additional smoothing of the data be performed after image reconstruction.

The EM algorithm is only guaranteed to converge to a local maximum, and the complexity of the data may imply multiple local maxima. There are stochastic variations of the EM algorithm other than the one discussed here, which are of possible interest, such as the Stochastic EM algorithm (Celeux and Diebolt 1985) and Stochastic Approximation EM (Delyon, Lavielle, and Moulines 1999). The basic idea of the stochastic variation is to inject random noise into the deterministic update of EM in the hope that the noise will "push" the method away from a local trap and hence lead to a better solution. Some of our preliminary work on the Stochastic EM algorithm suggests similar performance. Although the stochastic variation alleviates some of the trapping, it does not necessarily find all local maxima.

We have focused on model selection, and ignored the uncertainty in this selection. It is possible that a single best model does not exist. Therefore, model averaging is a direction worth exploring. Buckland, Burnham and Augustin (1997) suggested several ad hoc non-Bayesian approaches to account for model uncertainty. The smoothed AIC estimator is later embraced by Burnham and Anderson (2002), Claeskens and Hjort (2003), and Hjort and Claeskens (2003). It essentially constructs a weighted average of parameters of interest across candidate models, where the weight is proportional to the exponent of AIC, i.e. $w_m \propto \exp\{-\text{AIC}_m/2\}$. In our application, we have done some preliminary work in this direction. We first compute the estimated intensities $\mu_{z_i}^{\text{est.}}$ for each sub-model indexed by M. We then apply the above weight to obtain a weighted average across a series of sub-models. However, in our applications, the AIC for the best model is considerably smaller than that for competing models, so model averaging does not make a practical difference. As alternatives to the deterministic nature of the EM algorithm, there are stochastic approaches for similar models using evolutionary algorithms. Destrempes, Mignotte, and Angers (2005) extended the earlier

work by Francois (2002) using the Exploration/Selection/Estimation procedure, although we have not explored these alternatives.

We are currently working on a parallel Bayesian analysis of our proposed image model where the distribution of the number of hidden labels is estimated via reversible jump MCMC (Green, 1995). In the Bayesian framework, model averaging is a natural, and often argued, method of parameter estimation as model uncertainty is accounted for in the estimates.

In this manuscript we have concentrated in the change of contrast uptake from baseline to the three week visit, identified as a key visit. The data actually consist of a baseline image study and five follow-up image studies. Thus, modeling both the spatial and temporal aspects of the study is also of great interest.

Acknowledgements

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

Akaike H (1973), Information Theory and an Extension of the Maximum Likelihood Principle, *Second International Symposium on Information Theory*, 267.

Besag JE (1974), Spatial Interaction and the Statistical Analysis of Lattice Systems, *Journal of the Royal Statistical Society. Series B* **36**,192-236.

Besag JE (1986), On the Statistical Analysis of Dirty Pictures, *Journal of the Royal Statistical Society. Series B* **48**, 259-302.

Buckland ST, Burnham KP, and Augustin NH (1997), Model Selection: An Integral Part of Inference, *Biometrics* **53**, 603-618.

Burnham KP and Anderson DR (2002), Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, *Springer*.

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**, 4127-4136.

Celeux G and Diebolt J (1985), The SEM Algorithm: a Probabilistic Teacher Algorithm Derived from the EM Algorithm for the Mixture Problem, *Computational Statistics Quarterly* **2**, 73-82.

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

Claeskens G and Hjort NL (2003), The Focussed Information Criterion, *Journal of the American Statistical Association* **98**, 900-916.

Delyon B, Lavielle M, and Moulines E (1999), Convergence of a Stochastic Approximation Version of the EM Algorithm, *The Annals of Statistics* **27**, 94-128.

Dempster AP, Laird NM, and Rubin DB (1977), Maximum Likelihood from Incomplete Data vgia the EM Algorithm, *Journal of Royal Statistical Society, Series B* **39**, 1-38.

Deng H and Clausi DA (2004), Unsupervised Image Segmentation Using a Simple MRF Model with a New Implementation Scheme, *Pattern recognition* **37**, 2323-2335.

Destrempes F, Mignotte M, and Angers JF (2005), A Stochastic Method for Bayesian Estimation of Hidden Markov Random Field Models with Application to a Color Model, *Institute of Electrical and Electronics Engineers Transaction on Image Processing* **14**, 1096-1108.

Francois O (2002), Global Optimization with Exploration/Selection Algorithms and Simulated Annealing, *Annals of Applied Probability* **12**, 248-271.

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, *Proceedings of the National Academy of Sciences* **46**, 16759-16764;

Higdon D (1998), Auxiliary Variable Methods for Markov Chain Monte Carlo with Applications, *Journal of the American Statistical Association* **93**, 585-595.

Lakshmanan S and Derin H (1989), Simultaneous Parameter Estimation and Segmentation of Gibbs Ramdom Fields Using Simulated Annealing, *Institute of Electrical and Electronics Engineers Transactions on Pattern Analysis and Machnes Intelligence* **2**, 799-813.

Little RJA and Rubin DB (1983), On Jointly Estimating Parameters and Missing Data by Maximizing the Complete-data Likelihood, *American Statistician* **37**, 218-220.

Lei T and Udupa JK (2003), Image Analysis: Recovering the True Scene from Noisy Images, *Proceedings of the SPIE, Image Processing* **5032**, 1719-1727.

Louis T (1982), Finding the Observed Information Matrix when Using the EM Algorithm, *Journal of the Royal Statistical Society. Series B* **44**, 226-233.

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.

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, *Proceedings of the National Academy of Sciences* **102**, 5524-5529.

Panjwani DK and Healey G (1995), Markov Random Field Models for Unsupervised Segmentation of Textures Color Images, *Institute of Electrical and Electronics Engineers*

Transaction on Pattern Analysis and Machines Intelligence **17**, 939-954.

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

Rubin DB (1987), Multiple Imputation for Nonresponse in Sample Surveys, *Wiley*.

Schwarz G (1978), Estimating the Dimension of a Model, *Annals of Statistics* **6**, 461-464.

Sengur A, Turkoglu I, and Ince CM (2006), Unsupervised Image Segmentation Using Markov Random Fields, *Lecture Notes in Artificial Intelligence* **3949**, 158-167.

Swendsen RH and Wang JS (1987), Nonuniversal Critical Dynamics in Monte Carlo Simulations, *Physical Review Letters* **58**, 86-88.

Wei GCG and Tanner MA (1990), A Monte Carlo Implementation of the EM Algorithm and the Poor Man's Data Augmentation Algorithms, *Journal of the American Statistical Association* **85**, 699-704.

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.

Worsley KJ and Friston KJ (1995), Analysis of fMRI Time-Series Revisited - Again, *NeuroImage* **2**, 173.

Zhang J, Modestino JW, and Langan DA (1994), Maximum-likelihood Parameter Estimation for Unsupervised Stochastic Model-based Image Segmentation, *Institute of Electrical and Electronics Engineers Transactions on Image Processing* **3**, 404-420.

Zhang Y, Brady M, and Smith S (2001), Segmentation of Brain MR Images Through a Hidden Markov Random Field Model and the Expectation-Maximization Algorithm, *Medical Imaging, Institute of Electrical and Electronics Engineers Transactions* **20**, 45-57.

Appendix

As stated in section 2, the normalizing constant

$$g(\beta) = \sum_{\mathbf{z}\in\mathcal{S}^N} \exp\left\{\beta \sum_{i\sim i'} \mathrm{I}(z_i = z_{i'})\right\}$$

can not be evaluated analytically. Therefore in what follows, we define a set of quantities to numerically approximate it.

Let w denote the number of neighboring pairs on the lattice, i.e.

$$w = |\{i \sim i' : i, i' = 1, 2, \cdots, N\}|,$$

where $|A|$ is the cardinality of a set A. We define

$$U(\mathbf{z} \mid \beta) = \sum_{i\sim i'} \mathrm{I}(z_i = z_{i'}).$$

It is clear that $0 \leqslant U(\mathbf{z} \mid \beta) \leqslant w$ for all $\mathbf{z}$. We further use $a_j (0 \leqslant j \leqslant w)$ to denote the number of configurations with j pairs of same-labeled neighbors, i.e. $a_j = |\{\mathbf{z} : U(\mathbf{z} \mid \beta) = j\}|$. By definition, we have $\sum_{j=0}^{w} a_j = M^N$. Hence, the normalizing constant can be rewritten as $g(\beta) = \sum_{j=0}^{w} a_j \exp(\beta j)$. Assume U is a random variable with probability mass function

$$\Pr(U = j|\beta) = \frac{a_j \exp(\beta j)}{\sum_{j'=0}^{w} a_{j'} \exp(\beta j')}, \text{ for } 0 \leqslant j \leqslant w.$$

The moment generating function (m.g.f.) of U is

$$M_U(t) = E_\beta e^{tU} = \frac{\sum_{j=0}^{w} a_j e^{tj} \exp(\beta j)}{\sum_{j'=0}^{w} a_{j'} \exp(\beta j')}$$

with rth derivative

$$\frac{d^r M_U(t)}{dt^r} = \frac{\sum_{j=0}^{w} j^r a_j e^{tj} \exp(\beta j)}{\sum_{j'=0}^{w} a_j \exp(\beta j')}.$$

We claim the rth cumulant of U, k_r, is the rth derivative of $\log g(\beta)$, i.e. $k_r = \frac{d^r}{d\beta^r} \log g(\beta)$.

Specifically, when $r = 1, 2, 3$ we have

$$\begin{aligned}
k_1 &= \left.\frac{M_U'(t)}{M_U(t)}\right|_{t=0} = \left.\frac{\sum_{j=0}^{w} j a_j e^{tj} \exp(\beta j)}{\sum_{j'=0}^{w} a_{j'} e^{tj'} \exp(\beta j')}\right|_{t=0} = \frac{\sum_{j=0}^{w} j a_j \exp(\beta j)}{\sum_{j'=0}^{w} a_{j'} \exp(\beta j')} = \frac{g'(\beta)}{g(\beta)} = \frac{d}{d\beta} \log g(\beta) \\
k_2 &= \left.\frac{M_U(t) M_U''(t) - \{M_U'(t)\}^2}{\{M_U(t)\}^2}\right|_{t=0} \\
&= \left.\frac{\sum_{j=0}^{w} a_j e^{tj} \exp(\beta j) \sum_{j=0}^{w} j^2 a_j e^{tj} \exp(\beta j) - \{\sum_{j=0}^{w} j a_j e^{tj} \exp(\beta j)\}^2}{\{\sum_{j'=0}^{w} a_{j'} e^{tj'} \exp(\beta j')\}^2}\right|_{t=0} \\
&= \frac{\sum_{j=0}^{w} j^2 a_j \exp(\beta j) \sum_{j=0}^{w} a_j \exp(\beta j) - \{\sum_{j=0}^{w} j a_j \exp(\beta j)\}^2}{\{\sum_{j'=0}^{w} a_{j'} \exp(\beta j')\}^2} \\
&= \frac{g(\beta) g''(\beta) - \{g'(\beta)\}^2}{\{g(\beta)\}^2} = \frac{d^2}{d\beta^2} \log g(\beta) \\
k_3 &= \left.\frac{\{M_U(t)\}^2 M_U'''(t) - 3 M_U(t) M_U'(t) M_U''(t) + 2\{M_U'(t)\}^3}{\{M_U(t)\}^3}\right|_{t=0} \\
&= \frac{\{g(\beta)\}^2 g'''(\beta) + 3 g(\beta) g'(\beta) g''(\beta) - 2\{g'(\beta)\}^3}{\{g(\beta)\}^3} = \frac{d^3}{d\beta^3} \log g(\beta)
\end{aligned}$$

and so forth.

The above property provides the basis for approximating $\frac{g'(\beta)}{g(\beta)} = \frac{d}{d\beta} \log g(\beta)$ via a Taylor expansion around β_0:

$$\begin{aligned}
\frac{d}{d\beta} \log g(\beta) &\approx \frac{d}{d\beta} \log g(\beta_0) + (\beta - \beta_0) \frac{d^2}{d\beta^2} \log g(\beta_0) + \cdots + \frac{1}{r!} (\beta - \beta_0)^r \frac{d^{r+1}}{d\beta^{r+1}} \log g(\beta_0) \\
&= k_1 + (\beta - \beta_0) k_2 + \cdots + \frac{1}{r!} (\beta - \beta_0)^r k_{r+1}.
\end{aligned}$$

By pre-calculating $g(\beta)$ on a set of grid points for a range of M, we are able to interpolate the normalizing constant on non-grid points. Because $\frac{d}{d\beta}(\frac{g'(\beta)}{g(\beta)}) = \frac{d^2}{d\beta^2} \log g(\beta) = k_2$ is the second cumulant (variance) of a random variable U, which is always nonnegative, the above property also suggests the ratio $g'(\beta)/g(\beta)$ is a monotonically increasing function of β.

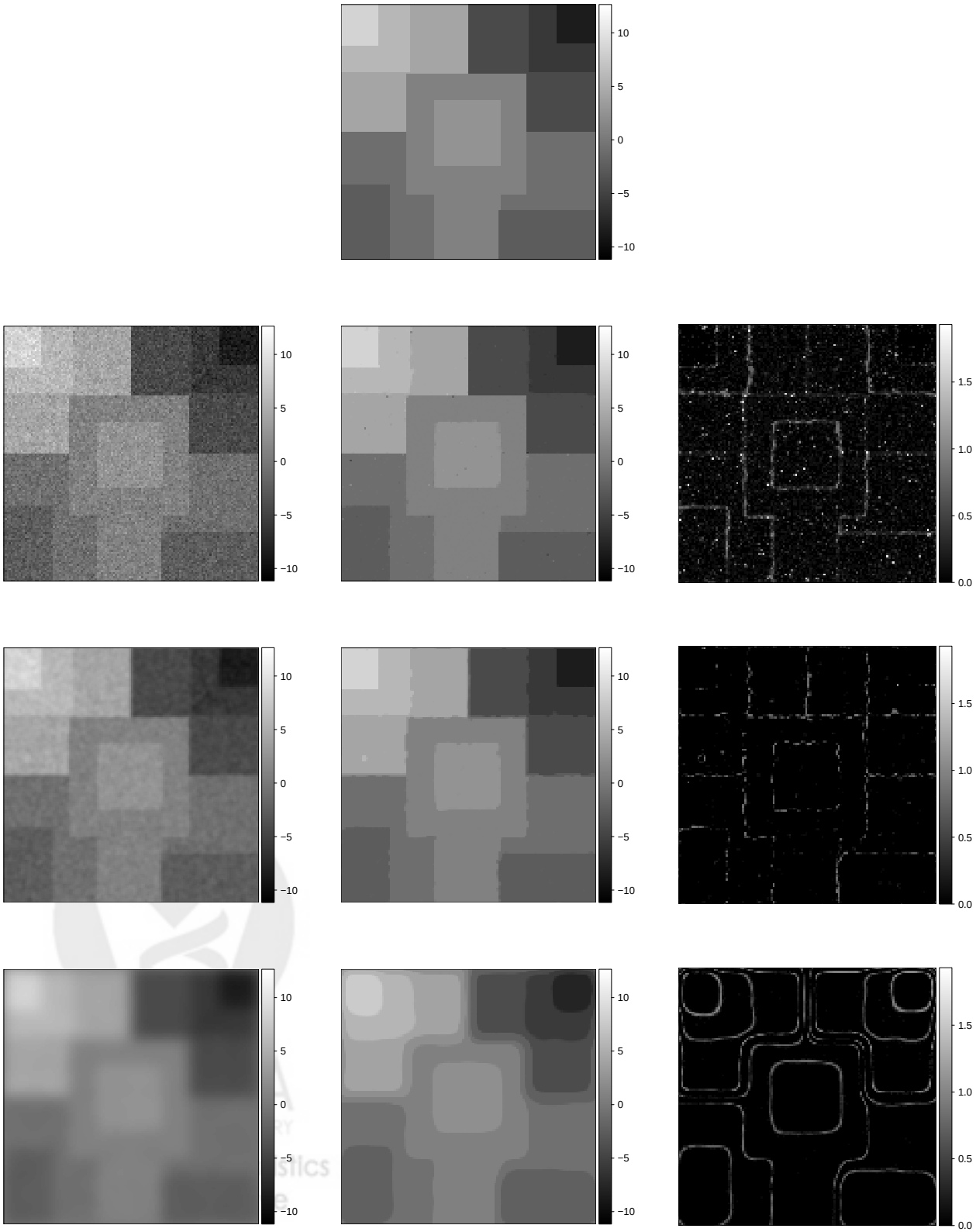

Figure 1. Simulation studies: the "true" scene (top), the observed images (left), the estimated pixel intensities (middle), and the associated standard deviation (right) when FWHM $= 0, 2$ and 8 from the second to fourth row ($M = 10$).

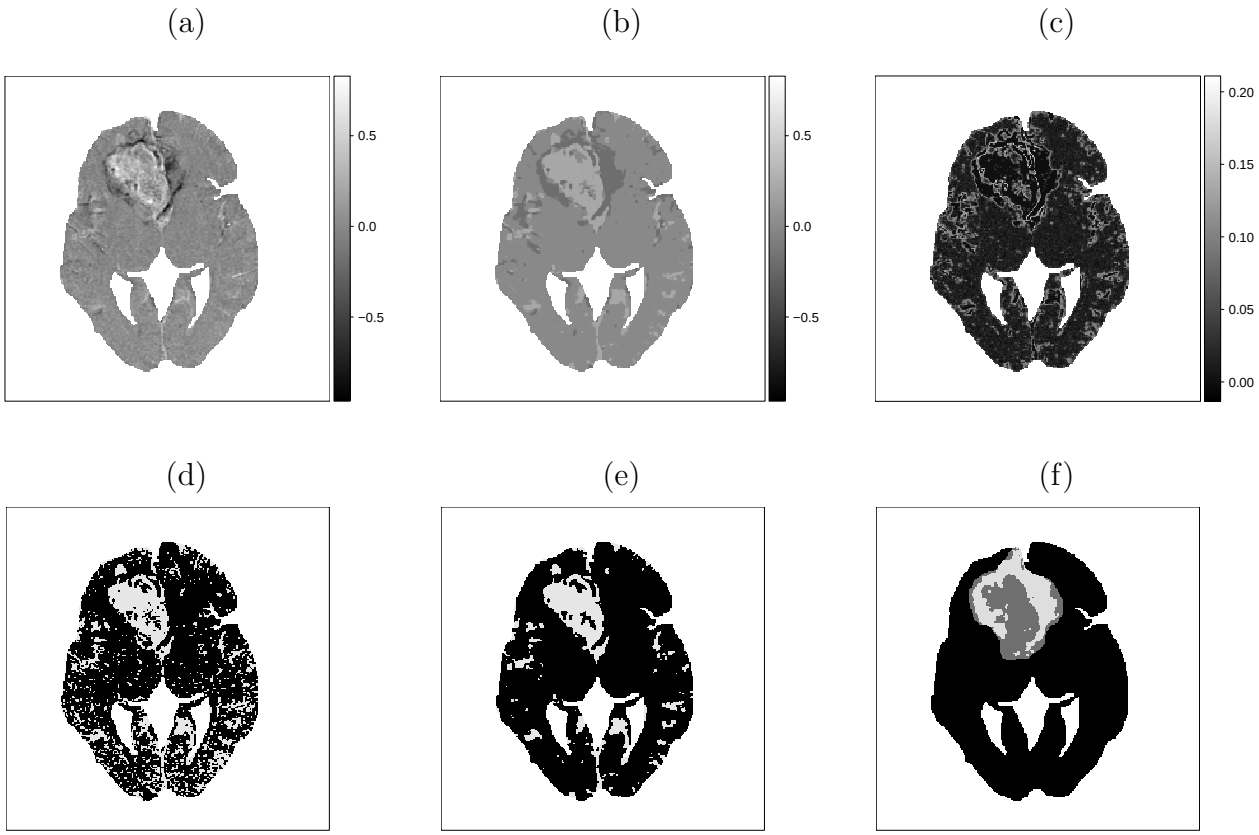

Figure 2. Results on patient 1: (a) observed change in contrast uptake (light shades standing for large increase); (b) estimated intensity $\mu_{z_i}^{\text{est.}} = E(\hat{\mu}_{Z_i} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_{M=3})$ and (c) standard deviation $SD(\hat{\mu}_{Z_i} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_{M=3})$; thresholded image (d) with and (e) without consideration of spatial structure; (f) baseline enhanced and non-enhanced tumor regions.

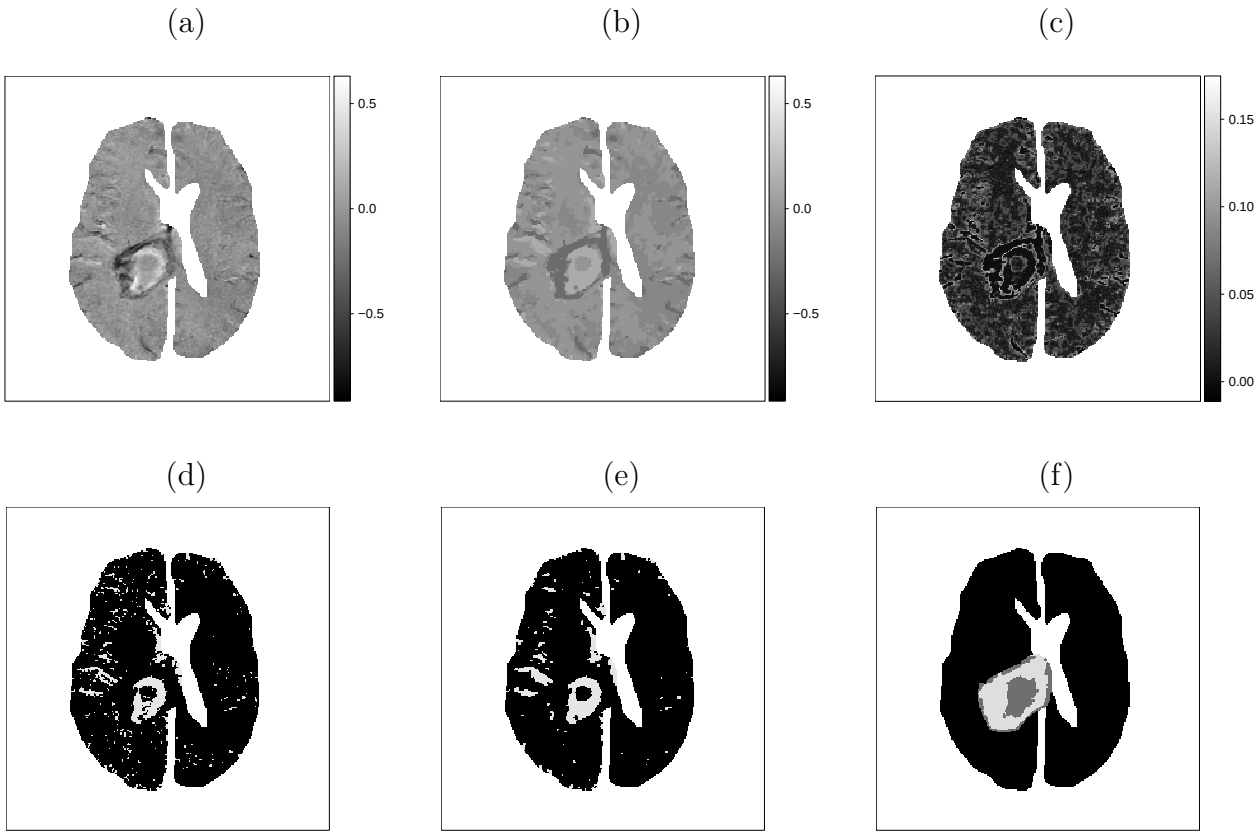

Figure 3. Results on patient 2: (a) observed change in contrast uptake (light shades standing for large increase); (b) estimated intensity $\mu_{z_i}^{\text{est.}} = E(\hat{\mu}_{Z_i} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_{M=4})$ and (c) standard deviation $SD(\hat{\mu}_{Z_i} \mid \mathbf{y}, \hat{\boldsymbol{\theta}}_{M=4})$; thresholded image (d) with and (e) without consideration of spatial structure; (f) baseline enhanced and non-enhanced tumor regions.

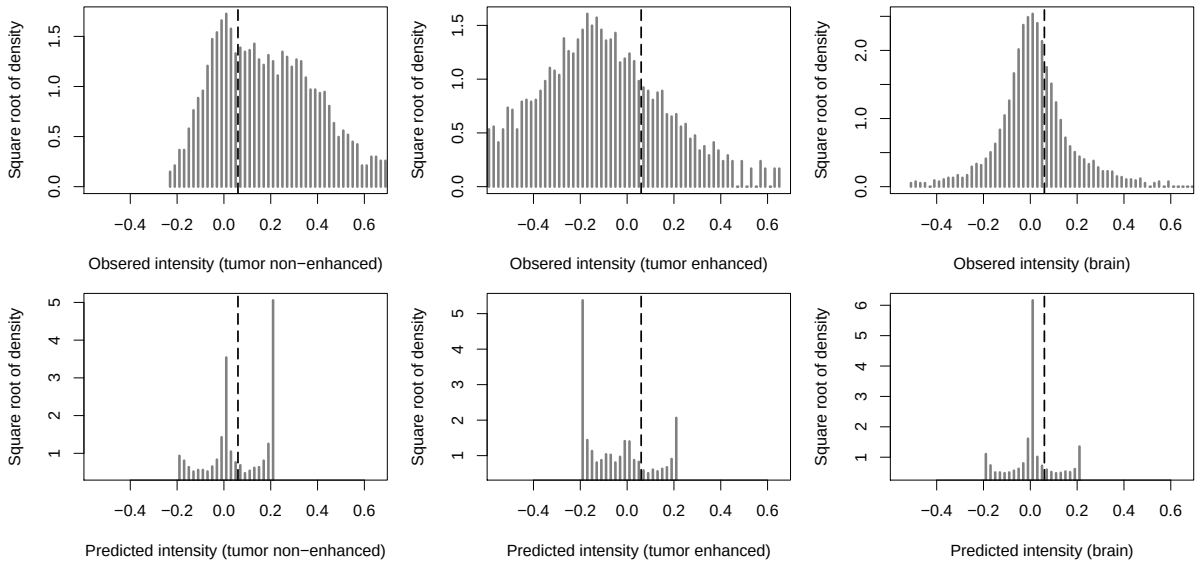

Figure 4. Histogram of observed (top) versus estimated (bottom) change in contrast uptake in the healthy tissue (left), initially non-enhanced (middle) and enhanced (right) tumor regions (patient 1).

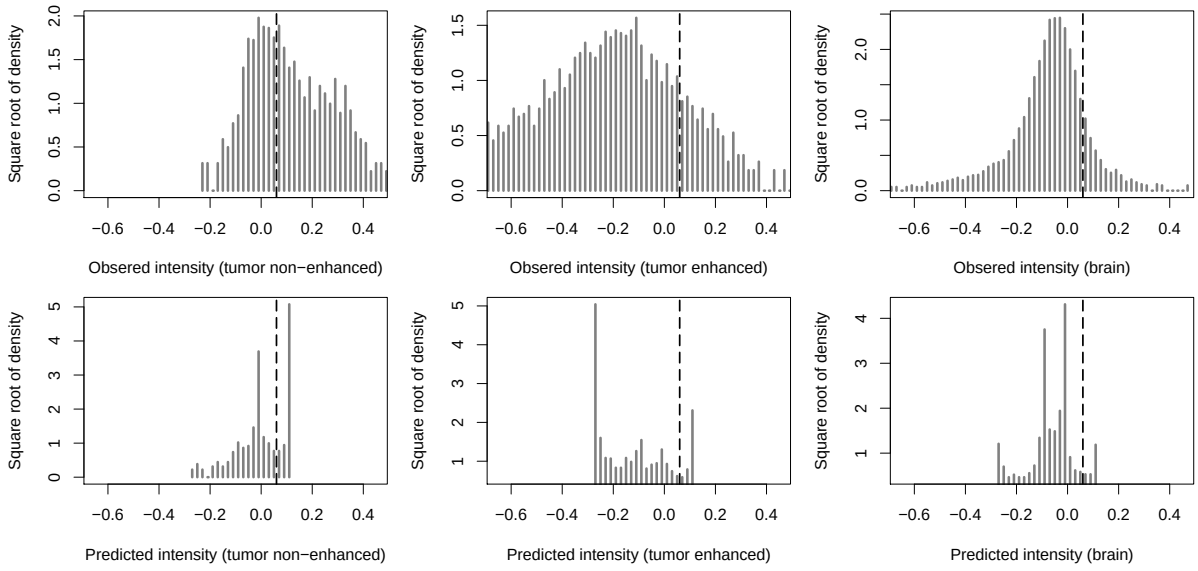

Figure 5. Histogram of observed (top) versus estimated (bottom) change in contrast uptake in the healthy tissue (left), initially non-enhanced (middle) and enhanced (right) tumor regions (patient 2).

Table 1
True and estimated parameter values under conditionally independent noise.

Comp.	Mean				Standard Deviation			
label(k)	μ_k	$\overline{\mu}_k$[a]	$\tilde{\mu}_k$	$\hat{\mu}_k(SD \times 10^2)$	σ_k	$\overline{\sigma}_k$	$\tilde{\sigma}_k$	$\hat{\sigma}_k(SD \times 10^2)$
1	-8.50	-5.86	-8.56	-8.56 (4.77)	1.00	2.09	0.94	0.94 (6.44)
2	-5.95	-4.34	-6.00	-5.94 (3.79)	1.00	1.04	0.96	1.00 (5.43)
3	-4.25	-2.89	-4.31	-4.26 (2.19)	1.00	1.04	0.97	0.98 (3.10)
4	-2.55	-1.68	-2.56	-2.53 (2.19)	1.00	0.88	0.98	0.99 (3.11)
5	-0.85	-0.79	-0.85	-0.85 (1.81)	1.00	0.76	0.98	0.99 (2.63)
6	0.85	0.25	0.84	0.84 (1.72)	1.00	0.64	0.99	1.00 (2.44)
7	2.55	1.21	2.58	2.55 (3.02)	1.00	0.79	0.95	0.96 (4.16)
8	4.25	2.19	4.31	4.24 (2.29)	1.00	1.08	0.99	1.00 (3.30)
9	5.95	4.29	6.04	5.96 (3.44)	1.00	0.94	0.91	0.94 (4.55)
10	8.50	6.03	8.53	8.50 (5.29)	1.00	1.97	1.01	1.05 (7.88)

[a] $\overline{\mu}_k$ and $\overline{\sigma}_k$ are estimates via a finite mixture model ignoring the spatial structure in the data, $\tilde{\mu}_k$ and $\tilde{\sigma}_k$ are estimates via Zhang et al. (2001) method, while $\hat{\mu}_k$ and $\hat{\sigma}_k$ are estimates via the proposed algorithm.

Table 2

Observed and estimated percentage of pixels in the initially enhanced and non-enhanced tumor region under various thresholds (90/95/97.5th percentile of healthy tissue contrast uptake before radiotherapy) with and without consideration of the spatial structure.

Percentile of healthy tissue at baseline	Patient 1				Patient 2			
	Enhanced(%)		Non-enhanced(%)		Enhanced(%)		Non-enhanced(%)	
	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.	Obs.	Pred.
90^{th}	17.1	16.9	64.2	60.1	11.5	14.0	53.2	55.8
95^{th}	14.7	14.4	62.1	59.1	10.2	12.6	51.6	54.5
97.5^{th}	12.5	12.5	59.8	56.8	9.4	11.5	49.9	53.1