Longitudinal image analysis of tumour–healthy brain change in contrast uptake induced by radiation

Xiaoxi Zhang

Pfizer, New York, USA

and Timothy D. Johnson, Roderick J. A. Little and Yue Cao

University of Michigan, Ann Arbor, USA

Summary. The work is motivated by a quantitative magnetic resonance imaging study of the differential tumour–healthy tissue change in contrast uptake induced by radiation. The goal is to determine the time in which there is maximal contrast uptake (a surrogate for permeability) in the tumour relative to healthy tissue. A notable feature of the data is its spatial heterogeneity. Zhang and co-workers have discussed two parallel approaches to 'denoise' a single image of change in contrast uptake from baseline to one follow-up visit of interest. In this work we extend the image model to explore the longitudinal profile of the tumour–healthy tissue contrast uptake in multiple images over time. We fit a two-stage model. First, we propose a longitudinal image model for each subject. This model simultaneously accounts for the spatial and temporal correlation and denoises the observed images by borrowing strength both across neighbouring pixels and over time. We propose to use the Mann–Whitney U-statistic to summarize the tumour contrast uptake relative to healthy tissue. In the second stage, we fit a population model to the U-statistic and estimate when it achieves its maximum. Our initial findings suggest that the maximal contrast uptake of the tumour core relative to healthy tissue peaks around 3 weeks after initiation of radiotherapy, though this warrants further investigation.

Keywords: Mann–Whitney U-statistic; Markov random field; Population model; Quantitative magnetic resonance imaging; Reversible jump Markov chain Monte Carlo methods; Spatial–temporal model

1. Introduction

This work is motivated by a pilot imaging study investigating the effects of radiation therapy on the vascular permeability of high grade gliomas, which are a particularly virulent type of brain cancer. Despite advances in both radiotherapy and chemotherapy, the overall survival time of a glioma patient has not significantly increased from a median survival of 1 year from diagnosis (Curran *et al.*, 1993).

Researchers hypothesize that tumour vasculature damage due to radiation would disrupt the blood–tumour barrier (BTB), at least transiently, allowing larger molecules to cross the BTB. If true, this suggests that chemotherapy should begin during the time when the damage to the BTB is at its peak. For instance, if the damage to the BTB peaks early during radiotherapy, greater tumour control could be achieved by starting chemotherapy during this time, in contrast with the current practice of sequentially administering radiotherapy followed by chemotherapy.

Address for correspondence: Xiaoxi Zhang, Pfizer Inc., 235 East 42nd Street, MS 219/7/1, New York, NY 10017, USA.
E-mail: xiaoxi.zhang@pfizer.com

11 subjects, who had been diagnosed with high grade gliomas, participated in this study conducted at the Department of Radiation Oncology, University of Michigan School of Medicine. The subjects underwent standard-of-care radiation therapy over a course of 4–6 weeks. Six imaging sessions were scheduled for each subject. The first image was obtained approximately 1 week before the onset of radiotherapy. The other five images were acquired at weeks 1 and 3 during therapy, and at 1, 3 and 6 months post therapy. Two T1-weighted magnetic resonance images, one before and one after the injection of the contrast enhancement agent—gadolinium diethylenetriaminepentaacetic acid—were obtained at each examination. All images were registered to the baseline computed tomography image that was used for planning treatment. The molecular weight of the gadolinium compound is approximately that of many chemotherapeutic molecules. Therefore, the contrast uptake (the volume of the compound that crosses the BTB, which is quantified as the log-ratio of the contrast-enhanced T1-weighted magnetic resonance imaging (MRI) image to the non-enhanced T1-weighted MRI image, e.g. Fig. 1) is a good surrogate of vascular permeability and may represent the volume of chemotherapeutic drug that crosses the BTB. The contrast uptake image is an example of quantitative MRI (QMRI), as it quantitates a physiological process, in this case vascular permeability. This is in contrast with an anatomical MRI examination that only allows visualization of anatomical structures and abnormalities and whose values have no absolute quantitative interpretation: only relative meaning.

Despite a large body of literature on medical image analysis, especially functional MRI, proper statistical analysis of QMRI data is limited. Two important features of QMRI data are spatial correlation and heterogeneity. Although they seem to be discordant features, an examination of Fig. 1 reveals that contrast uptake in healthy tissue is quite homogeneous and we would expect that there is high spatial correlation (at least locally). The tumour, which is visible in the upper left quadrant, is much more heterogeneous with respect to its contrast uptake. This heterogeneity can, roughly, be divided into two relatively homogeneous regions: the core of the tumour and its periphery, or annulus. The core of a tumour is typically hypoxic (low oxygen content due to a lack of blood supply) and appears dark, indicating low contrast uptake (e.g. Fig. 1). Conversely, the annulus is typically rich in neovasculature (newly formed blood vessels that are typically disorganized and leaky) and appears lighter in shade, indicating high contrast uptake (e.g. Fig. 1). It is known that hypoxia is protective against damage due to radiation and chemotherapy. Therefore, the core is usually a source of tumour regrowth after therapy. The focus of this imaging study is on demonstrating that radiation therapy can transiently increase vascular permeability in the core of a tumour.

In many QMRI studies, spatial information is ignored (e.g. Cao *et al.* (2005), Moffat *et al.* (2005) and Hamstra *et al.* (2005)). Zhang *et al.* (2008, 2010) developed an appropriate statistical model (a Gaussian hidden Markov random-field (MRF) model) that smooths the QMRI data while respecting the heterogeneity within the tumour, as well as between the tumour and healthy tissue. They proposed both a maximum likelihood and a Bayesian estimation method for image smoothing with edge preservation. However, their work focused on a single image—the change in contrast uptake from baseline to week 3—and thus ignored the temporal aspects of the data. In this paper we extend their Bayesian approach to account for the temporal profile of the contrast uptake in addition to the spatial correlation of the data.

We summarize the 'true' contrast uptake of the tumour relative to healthy tissue at each visit with the Mann–Whitney *U*-statistic. We build population models treating the *U*-statistic as the response variable to investigate the population temporal profile of the tumour–healthy tissue contrast uptake. The best population model is chosen on the basis of the predictive distribution criterion that was proposed by Gelfand *et al.* (1992).
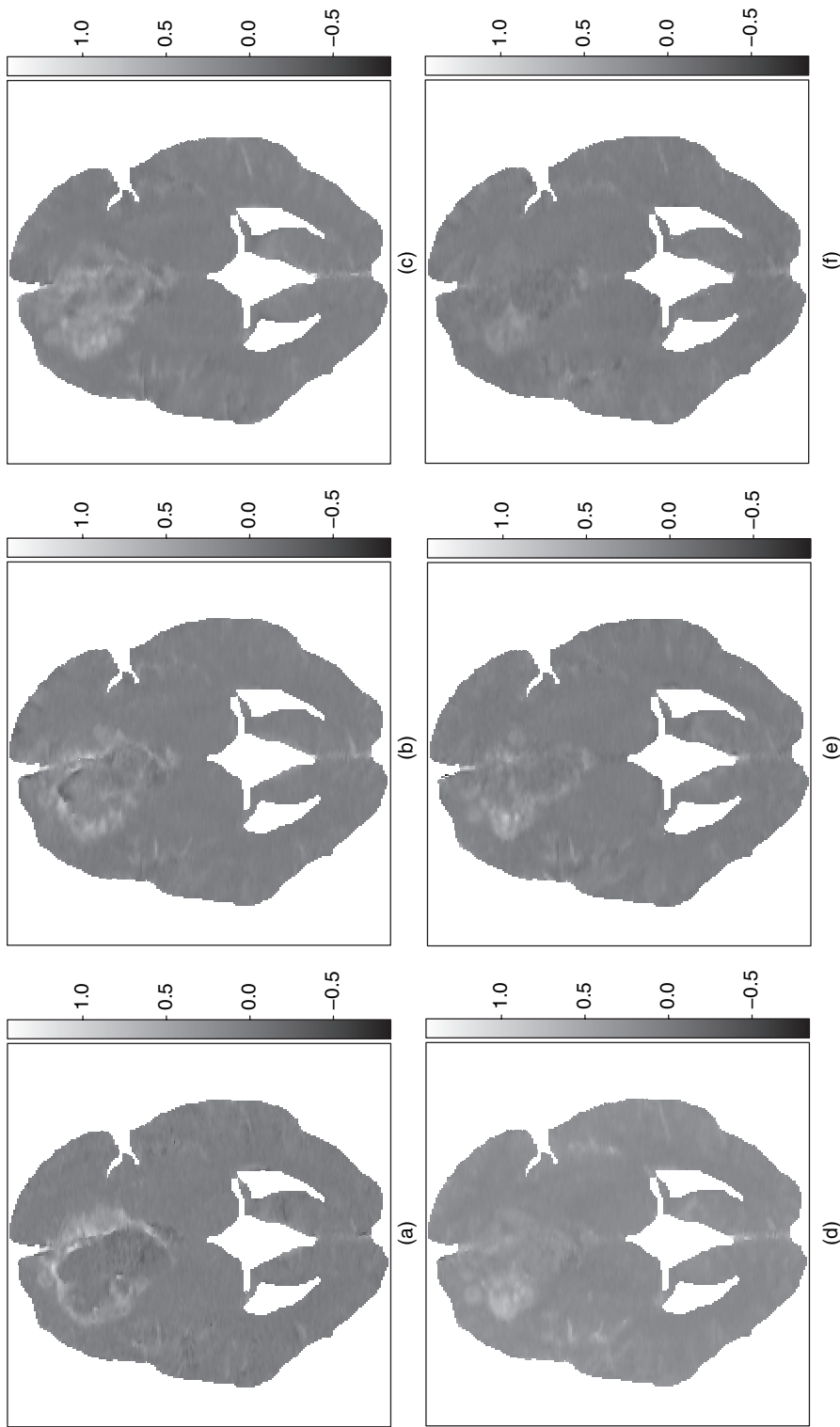
Fig. 1. Observed contrast uptake $\mathbf{y}_j$ for subject 1: (a) baseline $\mathbf{y}_1$; (b) 1 week after initiation of treatment, $\mathbf{y}_2$; (c) 3 weeks after initiation of treatment, $\mathbf{y}_3$; (d) 1 month after completion of treatment, $\mathbf{y}_4$; (e) 3 months after completion of treatment, $\mathbf{y}_5$; (f) 6 months after completion of treatment, $\mathbf{y}_6$

In the next section we describe the details of the model proposed. In Section 3 we outline its implementation. Results from simulation studies and the motivating study are presented in Section 4. A small sensitivity analysis is also conducted in the same section. We conclude the paper with a discussion of the model and ideas for future research.

2. Model specification

In this section, we first describe a longitudinal image model at the subject level that borrows strength both across pixels and over time. Later, we summarize the differential tumour–healthy tissue response to radiation by using the U-statistic, and then we build a population model treating the U-statistic as the outcome of interest.

2.1. Longitudinal image model

We build a longitudinal image model for each subject by using the following notation. Let j index the visits at time $t_1, t_2, \ldots, t_J$. Since we assume that the images are registered over time for each subject, we use a common index i ($i = 1, 2, \ldots, N$) to denote pixels regardless of visit (pixel i at time t_j and pixel i at time $t_{j'}$ are the same pixel). We denote the observed contrast uptake of pixel i at time t_j by y_{ij} and write the image at time t_j in vector form as $\mathbf{y}_j = (y_{1j}, \ldots, y_{Nj})^{\mathrm{T}}$ and all J images as $\mathbf{y} = (\mathbf{y}_1^{\mathrm{T}}, \ldots, \mathbf{y}_J^{\mathrm{T}})^{\mathrm{T}}$.

2.1.1. Distribution of the data

We decompose the observed contrast uptake into a true contrast uptake and a measurement error, i.e. $y_{ij} = \nu_{ij} + e_{ij}$ where $e_{ij} \sim N(0, \sigma_e^2)$, or $[y_{ij} | \nu_{ij}, \sigma_e^2] \sim N(\nu_{ij}, \sigma_e^2)$, for all i and j, as the normality assumption is justified in the MRI literature (e.g. Liang and Lauterbur (1999), chapter 8, and Lei and Udupa (2002)). The ν_{ij} denotes the unobserved true contrast uptake of pixel i at time t_j. We write in vector form $\boldsymbol{\nu}_j = (\nu_{1j}, \ldots, \nu_{Nj})^{\mathrm{T}}$ and $\boldsymbol{\nu} = (\boldsymbol{\nu}_1^{\mathrm{T}}, \ldots, \boldsymbol{\nu}_J^{\mathrm{T}})^{\mathrm{T}}$, the temporal and spatial structure of which is specified in the following section.

2.1.2. Prior distributions

(a) We specify a flat (improper) prior distribution for the standard deviation of the measurement error, i.e. $P(\sigma_e^2) \propto 1$.

(b) The decomposition of the observed contrast uptake into the 'truth' and conditionally independent noise in the above section is similar to that of Zhang *et al.* (2008, 2010). They imposed a spatial structure (an MRF prior; Besag (1974)) directly on the 'true' change in contrast uptake. Their model is sufficient when analysing a single image, i.e. the observed change in contrast uptake from baseline to week 3. However, we need additional modelling of the longitudinal perspective when analysing multiple QMRI images taken over time.

Since the true contrast uptake evolves over time during and after radiotherapy, we conceptualize the true contrast uptake at time t_j as the true contrast uptake from time t_{j-1} plus a change in contrast uptake and an evolution error. Here, 'evolution error' is in the sense that the true contrast uptake evolves over time with uncertainty. We borrow the terminology of measurement and evolution error from dynamic linear models (West and Harrison (1999), chapter 2) owing to the similarity of the structure of the proposed image model and dynamic linear model. For $j > 1$, this is formalized as $\nu_{ij} = \nu_{i,j-1} + \delta_{ij} + \varepsilon_{ij}$, $\varepsilon_{ij} \sim N(0, \tau_{ij}^2)$, where δ_{ij} is the structural 'change' (since $E(\boldsymbol{\nu}_j | \boldsymbol{\nu}_{j-1}, \boldsymbol{\delta}_j, \boldsymbol{\tau}_j^2) = \boldsymbol{\nu}_{j-1} + \boldsymbol{\delta}_j$)

and ε_{ij} is the evolution error of pixel i at time t_j. To keep the notation consistent, we define $\nu_{i0} \equiv 0$. Hence, $[\nu_{ij}|\nu_{i,j-1}, \delta_{ij}, \tau_{ij}^2] \sim N(\nu_{i,j-1} + \delta_{ij}, \tau_{ij}^2)$ for all i and j. Although δ_1 is the expectation of the true contrast uptake at the first visit (i.e. $E(\boldsymbol{\nu}_1|\boldsymbol{\nu}_0, \delta_1, \tau_1^2) = \delta_1$), we refer to δ_j as the true change in contrast uptake. We write in vector form $\delta_j = (\delta_{1j}, \ldots, \delta_{Nj})^{\mathrm{T}}$, $\delta = (\delta_1^{\mathrm{T}}, \ldots, \delta_J^{\mathrm{T}})^{\mathrm{T}}$, $\tau_j^2 = (\tau_{1j}^2, \ldots, \tau_{Nj}^2)^{\mathrm{T}}$ and $\tau^2 = (\tau_1^{2\mathrm{T}}, \ldots, \tau_J^{2\mathrm{T}})^{\mathrm{T}}$.

(c) Since radiation is administered in a gradient fashion, with the highest dose delivered to the tumour and tapering off in healthy tissue as a function of distance from the tumour, it is reasonable to assume that its effect (change in contrast uptake) is spatially correlated. Meanwhile, there is spatial heterogeneity within and surrounding the tumour. This motivates us to model the change in contrast uptake at each visit with an MRF prior, which not only accounts for the spatial correlation in the change but also respects the distinct boundaries between tumour and healthy tissue. Zhang *et al.* (2008, 2010) successfully applied MRF in modelling the observed change in contrast uptake $(y_{ij} - y_{ij'})$, which can be considered as the 'prototype' of the model proposed if we force $\varepsilon_{ij} \equiv 0$.

Specifically, let δ_{ij} $(i = 1, 2, \ldots, N_j)$ assume one of M_j possible values $(\mu_{1j}, \ldots, \mu_{M_j j})$ at time t_j. We write in vector form $\mu_j = (\mu_{1j}, \ldots, \mu_{M_j j})^{\mathrm{T}}$ and $\mu = (\mu_1^{\mathrm{T}}, \ldots, \mu_J^{\mathrm{T}})^{\mathrm{T}}$. We assign a uniform prior distribution for the M_j possible values: $\mu_{kj} \sim \mathrm{uniform}(\mu_{\min}, \mu_{\max})$ for all $1 \leqslant k \leqslant M_j$, where $\mu_{\min}$ and $\mu_{\max}$ are the minimum and maximum observed change in contrast uptake.

(d) Similarly to step (c), let τ_{ij}^2 $(i = 1, 2, \ldots, N_j)$ assume one of M_j possible values at time t_j, $(\sigma_{1j}^2, \ldots, \sigma_{M_j j}^2)$. We assume, *a priori*, $\sigma_{kj}^2 \sim \mathrm{Inv\text{-}gamma}(\alpha_\sigma, \beta_\sigma)$ for all $1 \leqslant k \leqslant M_j$. We write in vector form $\sigma_j^2 = (\sigma_{1j}^2, \ldots, \sigma_{M_j j}^2)^{\mathrm{T}}$, and $\sigma^2 = (\sigma_1^{2\mathrm{T}}, \ldots, \sigma_J^{2\mathrm{T}})^{\mathrm{T}}$. We set $\alpha_\sigma = 2.5$ and $\beta_\sigma = 5$, such that the prior mode of σ_{kj}^2 is 0.3. Later, in Section 4.1, we investigate the sensitivity of the results to other choices of α_σ.

(e) We introduce hidden MRF labels Z_{ij} on the state space $\mathcal{S}_j = \{1, \ldots, M_j\}$. Let Z_{ij} index the value of δ_{ij} and τ_{ij}^2 out of the M_j choices, i.e. $\delta_{ij} = \mu_{kj}$ and $\tau_{ij}^2 = \sigma_{kj}^2$ when $Z_{ij} = k$. The MRF prior on the hidden labels takes the form

$$P(\mathbf{Z}_j = \mathbf{z}_j|\beta_j, M_j) = g^{-1}(\beta_j, M_j) \exp(\beta_j \sum_{i \sim i'} I[z_{ij} = z_{i'j}]),$$

where $I[\cdot]$ is the indicator function, $i \sim i'$ denotes that pixels i and i' are neighbours (i.e. immediately adjacent) and

$$g^{-1}(\beta_j, M_j) = \sum_{\mathbf{Z}_j} \exp(\beta_j \sum_{i \sim i'} I[z_{ij} = z_{i'j}])$$

is the normalizing constant to be estimated via thermal integration (Ogata, 1989; Gelman and Meng, 1998). This model is also known as the Potts model in statistical physics (Potts, 1952). The collection of hidden labels, $\mathbf{z}_j = (Z_{1j} = z_{1j}, \ldots, Z_{Nj} = z_{Nj})^{\mathrm{T}}$, is called a configuration. We also write $\mathbf{z} = (\mathbf{z}_1^{\mathrm{T}}, \ldots, \mathbf{z}_J^{\mathrm{T}})^{\mathrm{T}}$. The spatial regularization parameter β_j controls the strength of the spatial correlation of the labels; for example, large β_j encourages smoother configurations.

(f) *A priori*, we assume that M_j is uniform on the integers from $M_{\min}$ to $M_{\max}$, where $M_{\min} = 2$ and $M_{\max} = 20$, such that the number of components is allowed to vary over a reasonable range. We write in vector form $\mathbf{M} = (M_1, \ldots, M_J)^{\mathrm{T}}$.

(g) Let $\beta_j \sim \mathrm{uniform}(0, \beta_{\max})$, where $\beta_{\max} = 3$. Similarly, we write in vector form $\beta = (\beta_1, \ldots, \beta_J)^{\mathrm{T}}$.

Note that we compute the ratio of the normalizing constant $g(\beta_j, M_j)$ off line on the grid values of $\beta_j = 0, 0.1, \ldots, 3.0$ and $M = 2, 3, \ldots, 20$ (appendix 1 of Zhang *et al.* (2010)). Its value on non-grid points is computed via linear interpolation. Furthermore, our model induces the

following covariance structure at each pixel over time. The covariance of the observed contrast uptake at the same pixel at any two visits (j_1 and j_2) is the variance of the true contrast uptake at the earlier visit ($j_{\min} = \min(j_1, j_2)$), i.e. $\mathrm{cov}(y_{ij_1}, y_{ij_2}) = \mathrm{var}(\nu_{ij_{\min}})$.

2.1.3. Joint distribution and conditional posterior distributions
Given the distribution of the data and the priors, the joint distribution is

$$P(\sigma_e^2) \prod_{j=1}^{J} \left(P(M_j)\, P(\beta_j)\, P(\mathbf{Z}_j | \beta_j, M_j) \prod_{k=1}^{M_j} \left[P(\mu_{kj})\, P(\sigma_{kj}^2) \prod_{i \in D_{kj}} \{ P(\nu_{ij} | \nu_{i,j-1}, \mu_{kj}, \sigma_{kj}^2)\, P(y_{ij} | \nu_{ij}, \sigma_e^2) \} \right] \right),$$

where $D_{kj} = \{i : Z_{ij} = k\}$ is the collection of pixels with the same label k at time t_j, i.e. a component. The conditional posterior distributions of μ_{kj}, σ_{kj}^2, ν_{ij} and σ_e^2 have closed form owing to conjugacy. By Bayes's theorem the conditional posterior distributions of $\mathbf{Z}_j$ and β_j satisfy

$$P(\mathbf{Z}_j | \cdot) \propto g^{-1}(\beta_j, M_j) \exp\left(\sum_{i \sim i'} \beta_j I[z_{ij} = z_{i'j}]\right) \prod_{k=1}^{M_j} \prod_{i \in D_{kj}} \sigma_{kj}^{-1} \exp\{-0.5 \sigma_{kj}^{-2} (\nu_{ij} - \nu_{i,j-1} - \mu_{kj})^2\},$$

$$P(\beta_j | \cdot) \propto g^{-1}(\beta_j, M_j) \exp\left(\sum_{i \sim i'} \beta_j I[z_{ij} = z_{i'j}]\right),$$

for $j = 1, \ldots, J$.

2.1.4. Marginal posterior mean of contrast uptake
Following Zhang *et al.* (2010), we characterize the contrast uptake by its posterior mean, i.e.

$$\nu_{ij} = \nu_{i,j-1} + \sum_{k=1}^{M_j} \mu_{kj} P(Z_{ij} = k | \mathbf{y}).$$

It can be estimated via Markov chain Monte Carlo draws $\delta_{ij}^{(t)}$ and $\nu_{i,j-1}^{(t)}$ at iteration t ($1 \leqslant t \leqslant T$): $\hat{\nu}_{ij} = T^{-1} \sum_{t=1}^{T} (\nu_{i,j-1}^{(t)} + \delta_{ij}^{(t)})$, where $\delta_{ij}^{(t)} = \mu_{kj}^{(t)}$ when $z_{ij}^{(t)} = k$. Other statistics of interest can be generated in similar ways.

2.2. Summarizing differential tumour–healthy tissue response
Although the longitudinal image model that was detailed in the above section estimates the true contrast uptake by smoothing the images, it does not directly answer the question of interest: when is there maximal contrast uptake in the tumour relative to healthy tissue? We, therefore, propose to use the Mann–Whitney U-statistic as a non-parametric and robust measure of separation in the true contrast uptake of tumour and healthy pixels. The U-statistic is interpreted as the probability that a randomly drawn value from the tumour (ν_T) is greater than a randomly drawn value from healthy tissue (ν_H). Compared with other non-parametric measures of differences between two sample distributions, e.g. the Kolmogorov–Smirnov statistic and Kullback–Leibler divergence, the U-statistic distinguishes $\nu_T > \nu_H$ from $\nu_H > \nu_T$. When the distributions are discrete, the U-statistic is particularly easy to compute: $U = \mathrm{Pr}(\nu_T > \nu_H) + 0.5\,\mathrm{Pr}(\nu_T = \nu_H)$. This is the method that is used in this paper. Large values of U suggest good separation of the two distributions. In the ideal case, when the two distributions are perfectly separated, U attains its maximum value of 1.

One benefit of using the U-statistic as a summary statistic of the differential tumour–healthy tissue contrast uptake is that a subject's healthy brain tissue serves as the control for the tumour tissue. As we have observed in some cases, the contrast uptake images from one visit can be overall brighter or darker than from another visit because of MRI machine receiver gain (e.g.

$\mathbf{y}_4$ in Fig. 1 is brighter than the rest of the visits). In such situations, the U-statistic is invariant, i.e. $P(\nu_T > \nu_H) = P(\nu_T + d > \nu_H + d)$ for an arbitrary shift in images intensities d.

The tumour core and annulus need to be examined separately because of their pathologic differences (Section 1). We divide each tumour into its core and annulus by applying the edge preserving smoothing that was discussed in Zhang *et al.* (2010) to the baseline contrast uptake image and then use the 95th percentile of the posterior mean (baseline contrast uptake) image as the threshold between tumour core and annulus. In principle, the threshold of tumour core *versus* annulus should be high (e.g. the 95th percentile of healthy tissue) to protect the bulk of healthy brain tissue. We are not aware of systematic or optimal choice of such a value. Therefore, we conduct a sensitivity analysis on our choice in Section 4.4.2.

2.3. Population model

In building the population model, we introduce a subject index h ($1 \leqslant h \leqslant H$). We denote the posterior mean of the U-statistic for subject h at time t_{hj} ($1 \leqslant j \leqslant J_h$) by A_{hj} (the grey profiles in Figs 2(a) and 2(c)). Since the U-statistic is bounded between 0 and 1, we apply a logit transformation to A_{hj} before fitting the population model, i.e. $\mathrm{logit}(A_{hj}) = \log(A_{hj}) - \log(1 - A_{hj})$.

The subject profiles in Figs 2(a) and 2(c) indicate the need for separate models for the tumour core and annulus. Considering the limited number of visits for each individual, we opt for a piecewise linear function of time (days since the initiation of radiation, t_{hj}) with a single variable change-point as the population mean model. The change-point location τ is treated as a model parameter. As considerable intersubject variability is seen in the profiles, we introduce individual random effects to capture the intersubject variability in the patient population, e.g. a random intercept, a random slope for the second linear segment or both. We define $[t]_+ = \max(0, t)$. Explicitly, the models M_1, M_2 and M_3 that we consider are respectively

$$\mathrm{logit}(A_{hj}) = \alpha + \beta_1 t_{hj} + \beta_2 [t_{hj} - \tau]_+ + b_{0h} + e_{hj},$$
$$\mathrm{logit}(A_{hj}) = \alpha + \beta_1 t_{hj} + (\beta_2 + b_{1h})[t_{hj} - \tau]_+ + e_{hj},$$
$$\mathrm{logit}(A_{hj}) = \alpha + \beta_1 t_{hj} + (\beta_2 + b_{1h})[t_{hj} - \tau]_+ + b_{0h} + e_{hj},$$

where $b_{0h} \sim N(0, \phi^2)$ and $b_{1h} \sim N(0, \psi^2)$ are the individual-specific random intercept and slope respectively, and $e_{hj} \sim N(0, \sigma^2)$ is the error term. We specify the following prior distributions for model parameters: $\phi^2 \sim \mathrm{Inv\text{-}gamma}(a_\phi, b_\phi)$, $\psi^2 \sim \mathrm{Inv\text{-}gamma}(a_\psi, b_\psi)$, $\sigma^2 \sim \mathrm{Inv\text{-}gamma}(a_\sigma, b_\sigma)$ and $\tau \sim \mathrm{uniform}(0, 183)$ and an improper flat prior density for $\boldsymbol{\theta} = (\alpha, \beta_1, \beta_2)^{\mathrm{T}}$, i.e. $P(\boldsymbol{\theta}) \propto 1$. These result in a proper joint posterior distribution. We set $a_\phi = a_\psi = a_\sigma = 1$, $b_\phi = b_\psi = 0.1$ and $b_\sigma = 0.5$ such that ϕ^2, ψ^2 and σ^2 are on reasonable scales, *a priori*.

We evaluate the adequacy of the population models via a cross-validation approach (Gelfand *et al.*, 1992), which aims to validate the conditional predictive distributions when each observation is left out. We denote the remaining data by $\mathbf{A}_{\backslash hj}$ when the jth visit of subject h is left out and we use a checking function $d_{hj} = \mathrm{logit}(A_{hj}) - E[\mathrm{logit}(\mathcal{A}_{hj})|\mathbf{A}_{\backslash hj}]$, where $E[\mathrm{logit}(\mathcal{A}_{hj})|\mathbf{A}_{\backslash hj}]$ is the posterior predictive mean of $\mathrm{logit}(\mathcal{A}_{hj})$.

The overall fit of a candidate model is evaluated as the sum of squares of the checking functions, i.e. $D = \Sigma_{h,j} d_{hj}^2$. The best-fitting model is the model with the smallest D (Gelfand *et al.*, 1992).

A straightforward approach is to fit the population model multiple times with one observation left out each time. Instead, we use the method that was described in Chen *et al.* (2000) to simplify the posterior predictive check. We denote the posterior distribution by $P(\boldsymbol{\theta}|\cdot)$ and
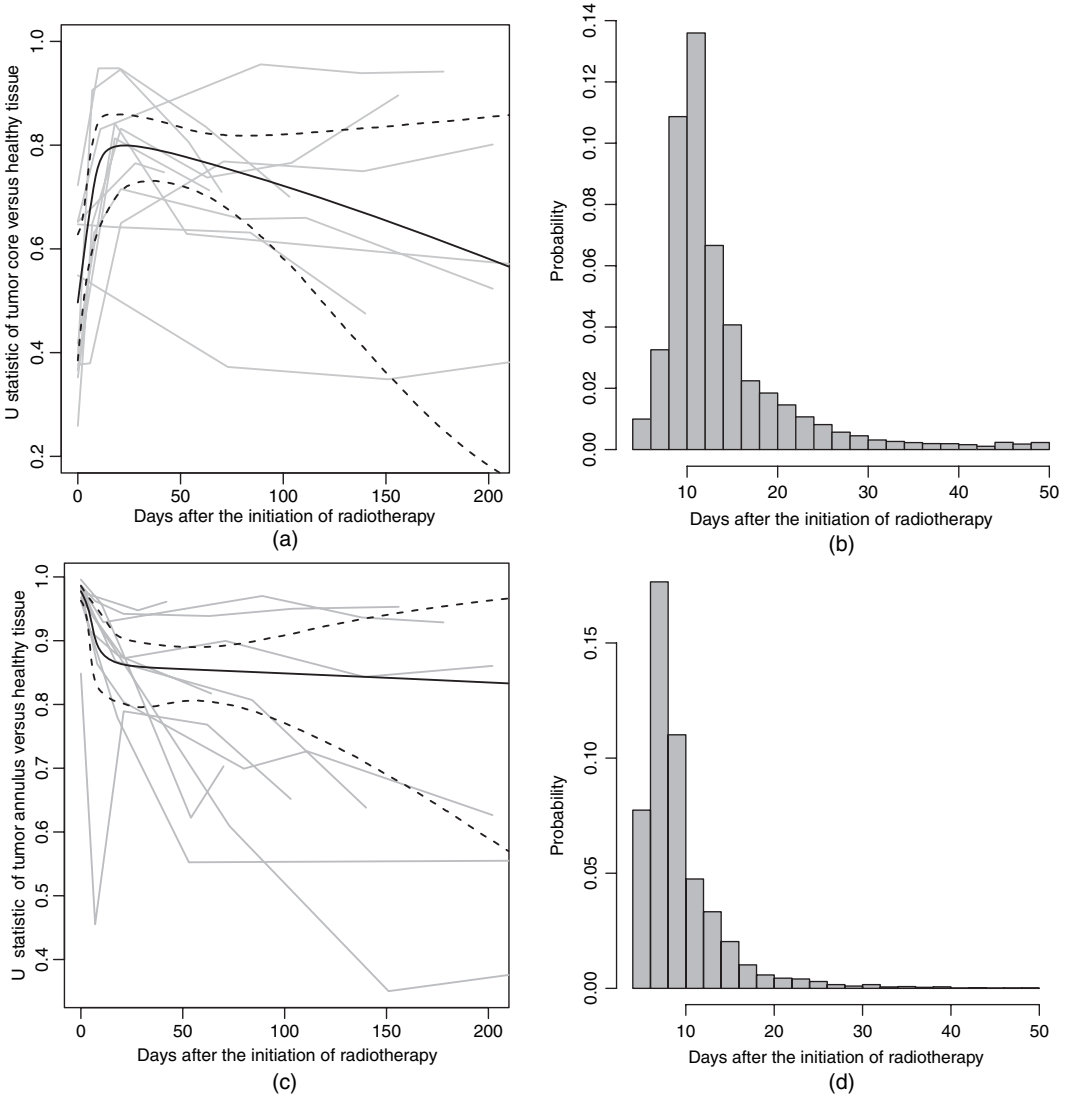
Fig. 2. Results on (a), (b) tumour core *versus* healthy tissue and (c), (d) tumour annulus *versus* healthy tissue: (a), (c) posterior mean of the *U*-statistic for all subjects and pointwise posterior mean and 95% credible interval of the population mean; (b), (d) posterior distribution of the change-point location τ

the distribution of the data by $P(A_{hj}|\boldsymbol{\theta})$. It follows from $P(\boldsymbol{\theta}|\mathbf{A}) \propto P(\boldsymbol{\theta}|\mathbf{A}_{\backslash hj}) P(A_{hj}|\boldsymbol{\theta})$, where $\mathbf{A} = (A_{hj}, \mathbf{A}_{\backslash hj})$, that

$$d_{hj} = \text{logit}(A_{hj}) - \int \int \text{logit}(\mathcal{A}_{hj}) P(\mathcal{A}_{hj}|\boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathbf{A}_{\backslash hj}) \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\mathcal{A}_{hj} \tag{1}$$

$$= \text{logit}(A_{hj}) - \frac{\int \int \text{logit}(\mathcal{A}_{hj}) P(\mathcal{A}_{hj}|\boldsymbol{\theta}) P(\boldsymbol{\theta}|\mathbf{A})/P(A_{hj}|\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta} \, \mathrm{d}\mathcal{A}_{hj}}{\int P(\boldsymbol{\theta}|\mathbf{A})/P(A_{hj}|\boldsymbol{\theta}) \, \mathrm{d}\boldsymbol{\theta}}.$$

This equation depends only on $P(\boldsymbol{\theta}|\mathbf{A})$ and not on $P(\boldsymbol{\theta}|\mathbf{A}_{hj})$. Therefore, the population model needs to be fitted only once with all the observations.

3.   Sketch of implementation

We implement the subject level longitudinal image model by running separate Monte Carlo chains for each subject. At every iteration, we update the parameters for the J_h visits of subject h sequentially. At time t_j, the conditional posterior distributions of $\boldsymbol{\mu}_j$, $\boldsymbol{\sigma}_j^2$ and $\boldsymbol{\nu}_j$ have closed form and are updated via standard Gibbs sampling steps. The spatial regularization parameter β_j requires a Metropolis–Hastings step. We use the Swendsen–Wang algorithm (Swendsen and Wang, 1987; Higdon, 1998; Zhang *et al.*, 2010) to update the hidden configurations. The most challenging part of the sampler is to update the number of components for each visit via reversible jump Markov chain Monte Carlo methods. The transdimensional proposal follows Zhang *et al.* (2010) with some modifications (see Appendix A for details). We then compute the empirical U-statistic for the tumour core and annulus separately on the basis of the posterior draws of $\boldsymbol{\nu}_{\mathrm{T}}$ and $\boldsymbol{\nu}_{\mathrm{H}}$ at that iteration. After all parameters indexed by visits have been updated, we finish the current iteration by drawing σ_e^2 via a Gibbs step and proceed to the next iteration.

We use the posterior mean of the U-statistic from all iterations after burn-in in the second-stage population model. The conditional posterior distributions of $\boldsymbol{\theta}$, ϕ^2, ψ^2 and σ^2 have closed form by Bayes's theorem and are updated via Gibbs steps. The change-point location τ is updated via a Metropolis–Hastings step. The best model is chosen via cross-validation (Gelfand *et al.*, 1992). The Markov chain Monte Carlo estimate of d_{hj} in equation (1) is

$$\hat{d}_{hj} = \mathrm{logit}(A_{hj}) - \frac{S^{-1}\sum\limits_{s=1}^{S}\mathrm{logit}(\mathcal{A}_{hj}^{(s)})/P(A_{hj}|\boldsymbol{\theta}^{(s)})}{S^{-1}\sum\limits_{s=1}^{S}1/P(A_{hj}|\boldsymbol{\theta}^{(s)})},$$

where $(\mathcal{A}^{(s)}, \boldsymbol{\theta}^{(s)})$ are posterior draws at iteration s $(1 \leqslant s \leqslant S)$. We fit models M_1–M_3 and select the model with the smallest $\hat{D} = \Sigma_{h,j}\hat{d}_{hj}^2$.

4.   Results

To evaluate the performance of the longitudinal image model, we conduct a series of simulation studies under various scenarios. For each scenario, we generate multiple simulations, indexed by $l = 1, 2, \ldots, L$. We compute the average mean-squared error aMSE and average bias abias for the true contrast uptake estimate for simulation l at time t_j, both averaged over all pixels in the same image:

$$\mathrm{aMSE}(\boldsymbol{\nu}_j^l) = N^{-1}\sum_{i=1}^{N}(\hat{\nu}_{ij}^l - \nu_{ij}^{l,\mathrm{true}})^2,$$

$$\mathrm{abias}(\boldsymbol{\nu}_j^l) = N^{-1}\sum_{i=1}^{N}(\hat{\nu}_{ij}^l - \nu_{ij}^{l,\mathrm{true}}),$$

where $\nu_{ij}^{l,\mathrm{true}}$ is the true contrast uptake that is used to generate the lth simulation and $\hat{\nu}_{ij}^l$ is its estimate (i.e. posterior mean). We also define the average aMSE and abias across the L simulations as $\overline{\mathrm{MSE}}(\boldsymbol{\nu}_j) = L^{-1}\Sigma_{l=1}^{L}\mathrm{aMSE}(\boldsymbol{\nu}_j^l)$ and $\overline{\mathrm{bias}}(\boldsymbol{\nu}_j) = L^{-1}\Sigma_{l=1}^{L}\mathrm{abias}(\boldsymbol{\nu}_j^l)$. To illustrate the benefits of modelling the temporal correlation, we compare this method with a set of cross-sectional models on the simulated data, i.e. fitting an independent model for the observed contrast

uptake image at each visit by following Zhang *et al.* (2010). Essentially, the cross-sectional model assumes that $y_{ij} = \eta_{ij} + e_{ij}$, $e_{ij} \sim N(0, \omega_{ij}^2)$, where the values of η_{ij} and ω_{ij}^2 are indexed by hidden label Z_{ij}, i.e. $\eta_{ij} = \xi_{kj}$ and $\omega_{ij}^2 = v_{kj}^2$ when $z_{ij} = k$. We then run the algorithm on the real data set and summarize the differential tumour–healthy tissue response with the U-statistic, and select the best-fitting population model for the tumour core and annulus separately.

4.1. Simulation study of random change images

We assume that there are six images ($J = 6$) on a 128×128 lattice. We divide the lattice into regions of various shapes (Fig. 3). The same skeleton is used for all visits. For each visit, we assign the same true change to all pixels in the same region ($\delta_{ij} = \delta_{i'j}$ for all i and i' in the same region at the same visit j), which is independently and randomly drawn from eight candidate levels, i.e. $M_j = 8$ and $\boldsymbol{\mu}_j = (-5.95, -4.25, -2.55, -0.85, 0.85, 2.55, 4.25, 5.95)^{\mathrm{T}}$ for all j. We set the standard deviation of all evolution errors to 0.7, i.e. $\sigma_{kj} = 0.7$ for all j and k. The standard deviation of the measurement errors is $\sigma_e = 1$. Figs 4(a)–4(c) shows an example of the true change images. We follow the model assumptions and generate the true contrast uptake images from $\nu_{ij} = \nu_{i,j-1} + \delta_{ij} + \varepsilon_{ij}$. The observed images are generated from $y_{ij} = \nu_{ij} + e_{ij}$. We generate 20 sets of observed images from 20 true change images in the above fashion ($L = 20$).

We run the algorithm proposed on each of the 20 simulated data sets. The pixelwise posterior means of $\boldsymbol{\delta}_j$ and $\boldsymbol{\nu}_j$ for the example data set are displayed in Figs 4(d)–4(f) and 4(j)–4(l). The posterior mean images of the true change are similar to the corresponding true values. However, we note that the errors are more likely to occur on the boundaries of two adjacent regions when they have means that are close to one another. By modelling the temporal correlation, the posterior mean images of the true contrast uptake ($\hat{\boldsymbol{\nu}}_j$) are less noisy than the observed images ($\mathbf{y}_j$) and the edges between regions with different true values are preserved.

We also compare the proposed longitudinal model with J independent cross-sectional models, one for each visit. The cross-sectional model assumes that each observed image is a hidden MRF deteriorated by Gaussian noise. At each time t_j, the true pixel mean η_{ij} is the counterpart of the true contrast uptake ν_{ij} in the model proposed. We compare aMSEs of the two approaches averaged over pixels within an image. Both methods have minimal bias. However, the longitudinal approach proposed produces uniformly smaller aMSEs than the cross-sectional model by borrowing strength over time. Because the noise (both evolution and measurement errors) adds up over visits, the mean-squared error is non-decreasing as the signal-to-noise ratio decreases
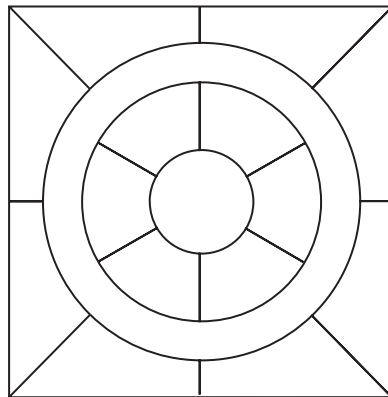

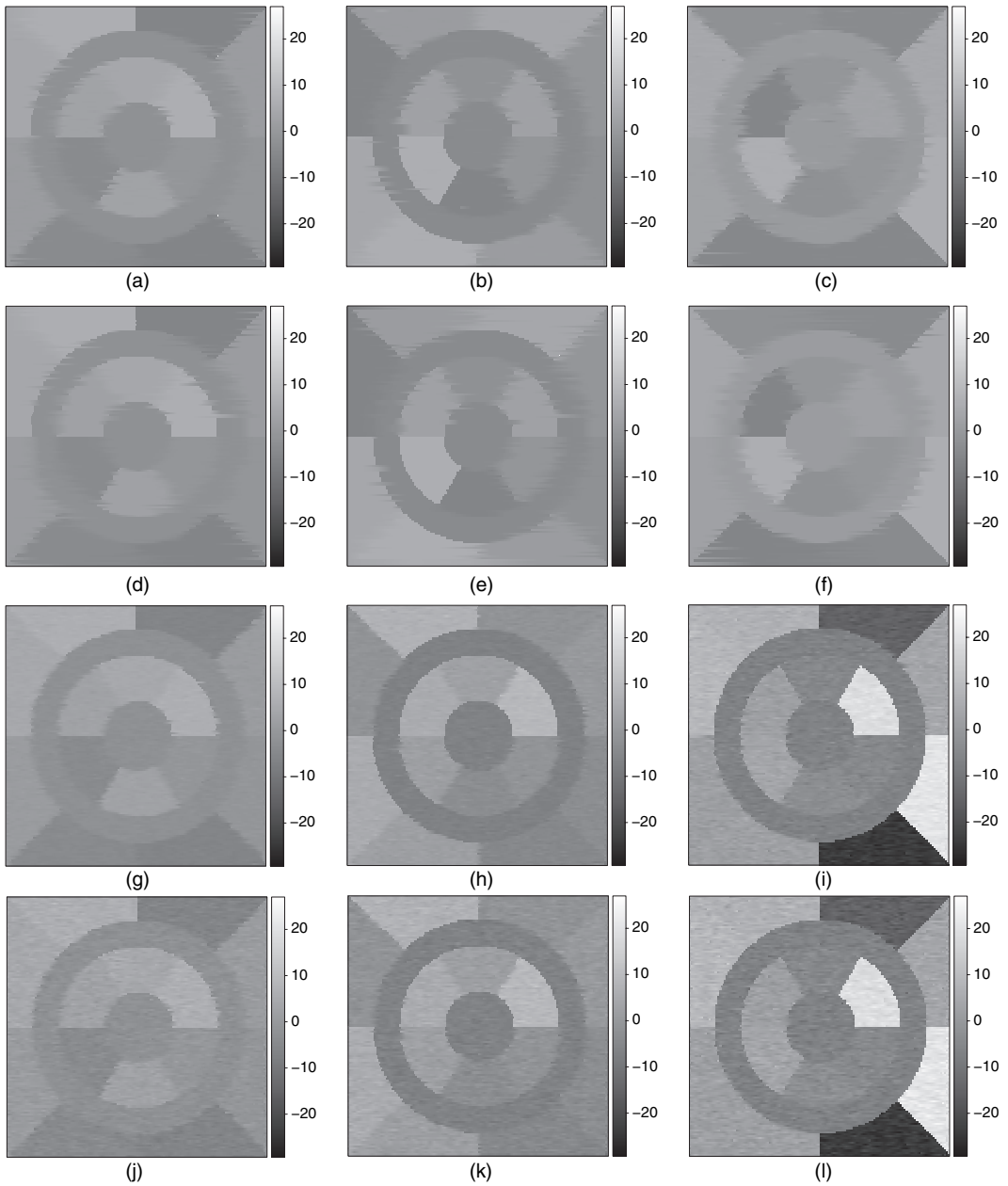**Fig. 3.**   Skeleton of true change image

Fig. 4. Simulation study of random true change images in contrast uptake δ_j, posterior mean $\hat{\delta}_j$, observed contrast uptake $\mathbf{y}_j$, and posterior mean of true contrast uptake $\hat{\nu}_j$: (a) δ_1; (b) δ_2; (c) δ_6; (d) $\hat{\delta}_1$; (e) $\hat{\delta}_2$; (f) $\hat{\delta}_6$; (g) $\mathbf{y}_1$; (h) $\mathbf{y}_2$; (i) $\mathbf{y}_6$; (j) $\hat{\nu}_1$; (k) $\hat{\nu}_2$; (l) $\hat{\nu}_6$

for both approaches. The aMSEs of the cross-sectional model increase roughly as multiples of its baseline value whereas the longitudinal model increases only moderately for the follow-up visits ($j > 1$) after borrowing strength across visits. By way of data generation, the true contrast uptake is essentially continuous. We argue that the posterior draws of ν from the longitudinal model are on a continuous scale for $j > 1$, whereas the posterior draws of η in the cross-sectional

model are discrete. This helps to explain why the method proposed fits the data better with smaller aMSEs.

We investigated the sensitivity of the results to alternative choices of the hyperprior, i.e. β_σ equal to 1 and 5 instead of 3. $\overline{\text{MSE}}(\nu_j)$ is not sensitive to the choice of λ—the change in aMSE is less than 5%.

4.2.   Simulation study of no treatment effect

We generate a second set of simulations by assuming no treatment effect. We assume that the regions within the outer ring are the 'tumour' with relatively low baseline contrast uptake, whereas the rest of the regions are 'healthy tissue'. At baseline t_1, we assign the tumour regions with means randomly drawn from $(-5.95, -4.25, -2.55, -0.85)$, and the means of the healthy regions are drawn from $(-2.55, -0.85, 0.85, 2.55, 4.25, 5.95)$, i.e. the tumour has lower values on average than the surrounding regions at baseline. The rest of the parameters remain the same as in the above section. We assume no structural change in contrast uptake for all subsequent visits, i.e. no treatment effect or $\delta_j = 0$ for $j > 1$. The ν_j and y_j are then generated following model assumptions. We generate 20 sets of such data sets and run the algorithm proposed on each set. The aMSEs of the proposed longitudinal model are significantly smaller than the MSEs of the cross-sectional model (Table 1) with the comparative advantage increasing over time.

4.3.   Simulation study of treatment effect

We simulate a third scenario assuming a treatment effect. At baseline, δ_1 is generated in the same way as the no-treatment-effect scenario. During the hypothetical radiotherapy period (t_2 and t_3), we assume that the tumour regions are more responsive to radiation than are the healthy regions—the structural changes in the tumour regions are higher on average than in the healthy regions, i.e. the mean values are randomly drawn from $(2.55, 4.25)$ and $(-0.85, 0.85)$ respectively. Hence, the structural change in the centre regions is brighter than the surrounding regions during the treatment period. Assuming no long-term treatment effect after its completion, we set $\delta_j = 0$ at times t_4, t_5 and t_6. The rest of the parameter values are the same as in Sections 4.1 and 4.2. Again, 20 sets of such simulations are generated ($L = 20$). The mean-squared errors of the longitudinal model proposed are once again smaller than the mean-squared errors of the cross-sectional model (Table 1) with the difference increasing over time.

Table 1.   aMSE and bias of the method proposed *versus* a cross-sectional approach—simulation studies of random-change images, no treatment effect and treatment effect

| *Effect*	*aMSEs* $(\times 10^{-1})$ *for the following values of j:*					
	j = 1	*j = 2*	*j = 3*	*j = 4*	*j = 5*	*j = 6*
Random						
Longitudinal	2.60 ± 0.03	3.28 ± 0.07	3.45 ± 0.05	3.54 ± 0.04	3.89 ± 0.08	5.20 ± 0.08
Cross-sectional	5.22 ± 0.47	9.81 ± 0.40	14.32 ± 0.73	19.90 ± 1.18	24.27 ± 1.09	29.2 ± 1.18
No treatment effect						
Longitudinal	3.04 ± 0.10	3.75 ± 0.05	3.74 ± 0.05	3.85 ± 0.04	3.92 ± 0.03	5.67 ± 0.07
Cross-sectional	5.64 ± 0.16	11.45 ± 0.79	17.01 ± 1.07	21.35 ± 0.95	25.59 ± 0.72	29.62 ± 0.78
Treatment effect						
Longitudinal	2.71 ± 0.15	3.48 ± 0.19	3.80 ± 0.07	3.83 ± 0.04	3.93 ± 0.03	5.62 ± 0.07
Cross-sectional	5.63 ± 0.16	9.83 ± 0.65	14.66 ± 0.81	20.30 ± 1.51	25.26 ± 1.70	30.36 ± 1.99
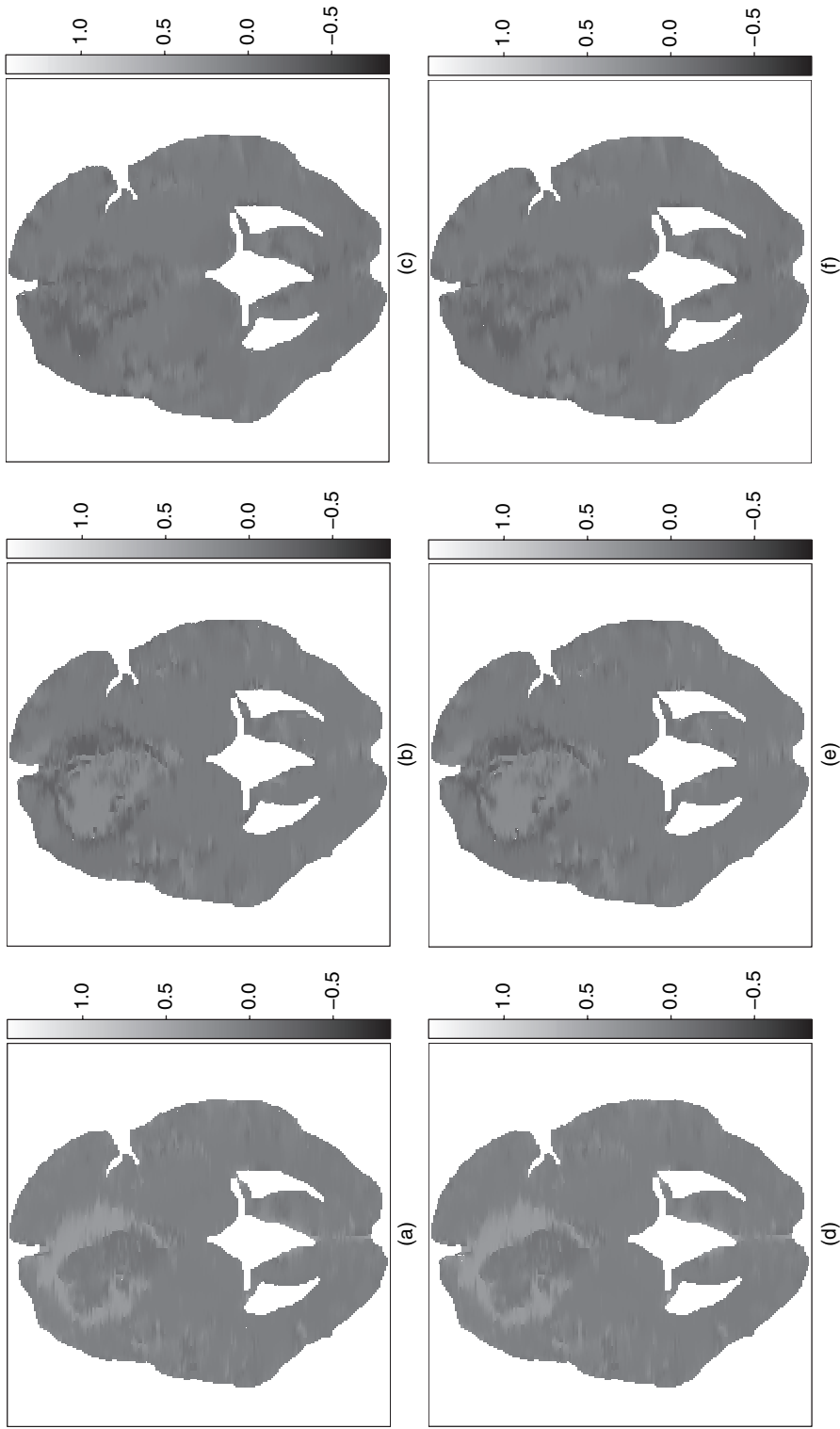
Fig. 5. Results for subject 1: posterior means of (a) $\hat{\delta}_1$, (b) $\hat{\delta}_2$ and (c) $\hat{\delta}_6$, and posterior means of (d) $\hat{\nu}_1$, (e) $\hat{\nu}_2$ and (f) $\hat{\nu}_6$

4.4.   Application

With satisfactory results from the simulation studies, we fit the real data with the image model proposed. Taking subject 1 for example, it takes about 4.5 h to complete 60 000 iterations on a Sun X2200 machine with a central processor unit speed of 2.6G Hz and 16 Gbytes of random-access memory. The posterior mean of true contrast uptake (Fig. 5) smooths the observed images (Fig. 1). The posterior mean image of true change in contrast uptake (Fig. 5) reveals a differential pattern for the diseased and healthy tissue.

4.4.1.   Population model

The tumour boundary was outlined by a radiation oncologist. We use the 95th percentile of baseline healthy tissue contrast uptake to segment the tumour core from the annulus. This roughly corresponds to the division of the tumour into enhanced (annulus) and non-enhanced (core) regions in the original analysis of Cao *et al.* (2005). We fit separate population models for them (see Fig. 2(a) for the core and Fig. 2(c) for the annulus). The population models are fitted after applying a logit transformation of the posterior means of the U-statistics, and then transformed back to its original scale as discussed in Section 2.2.

For the tumour core, the best-fitting model is M_3 with both a random intercept and a random slope for the second linear segment (Table 2). It profiles the steep increase in the population mean soon after the initiation of radiotherapy, and the moderate decrease afterwards. The change-point is around 10 days after the initiation of radiotherapy (the mode is 10 days, and the first and third quartiles are 9 and 14; Fig. 2(b)). The full and broken black curves in Fig. 2(a) display the posterior mean and its pointwise 95% credible interval of the population mean. The population mean reaches its maximum at around 3 weeks after the initiation of radiation. We see that the profile of one subject (the bottom subject in Fig. 2(a)) is notably different from the rest of the subjects. We refit the model excluding this subject. The parameter estimates do not change substantially (Table 2).

Model M_2 is the best-fitting model for the tumour annulus (Table 2). Fig. 2(c) shows that the population mean first decreases and then remains rather flat. The posterior distribution of the change-point location centres near 1 week after the initiation of radiotherapy (the mode is 7 days, and the first and third quartiles are 6 and 10; Fig. 2(d)). We also note that the profile of one subject fluctuates more than usual in Fig. 2(c) and refit the model excluding this subject; the parameter estimates are listed in Table 2. The change in the parameter values is small.

Assuming that contrast uptake is a reasonable surrogate of vascular permeability, and hence the delivery of chemotherapeutic agents, the findings suggest that chemotherapy should begin no later than 3 weeks after the initiation of radiotherapy. At this time, the relative exposure of the tumour core potentially reaches its maximum despite the slight decrease in the annulus. However, this warrants confirmation from more definitive studies, as the number of subjects in this pilot study is limited. Another implication is that the heterogeneity of tumour response to radiation may suggest a dynamic treatment modification during the course of radiotherapy, depending on the tumour's response.

4.4.2.   Sensitivity analysis

In the above analysis, the tumour core and annulus are segmented on the basis of the 95th percentile. We conduct a sensitivity analysis on other choices of threshold values of the tumour core *versus* annulus, i.e. the 90th and 97.5th percentiles of healthy tissue contrast uptake at baseline. The parameter estimates of the tumour core are given in Table 2. Using a higher threshold, a larger portion of the tumour with high baseline contrast uptake compared with other core

Table 2. Sensitivity analysis of posterior parameter estimates for the population model for tumour core and annulus *versus* healthy tissue

Parameter	90th percentile		95th percentile		97.5th percentile	
	All subjects	Excluding one	All subjects	Excluding one	All subjects	Excluding one
Core						
τ	12.6 ± 4.4	12.3 ± 3.5	13.7 ± 7.2	12.9 ± 5.2	16.2 ± 10.1	15.4 ± 8.8
α	-0.21 ± 0.21	-0.26 ± 0.19	-0.03 ± 0.24	-0.07 ± 0.22	0.15 ± 0.26	0.10 ± 0.25
β_1	0.14 ± 0.05	0.15 ± 0.04	0.13 ± 0.05	0.14 ± 0.05	0.11 ± 0.06	0.12 ± 0.06
β_2 $(\times 10^{-2})$	-0.65 ± 0.42	-0.75 ± 0.42	-0.60 ± 0.52	-0.70 ± 0.49	-0.61 ± 0.64	-0.74 ± 0.65
ϕ	0.69 ± 0.09	0.62 ± 0.08	0.75 ± 0.10	0.67 ± 0.09	0.78 ± 0.10	0.71 ± 0.09
ψ	0.14 ± 0.03	0.15 ± 0.04	0.14 ± 0.03	0.15 ± 0.04	0.14 ± 0.03	0.15 ± 0.04
σ	0.13 ± 0.03	0.14 ± 0.03	0.13 ± 0.03	0.14 ± 0.03	0.13 ± 0.03	0.14 ± 0.03
Annulus						
τ	14.1 ± 10.0	15.0 ± 5.3	10.1 ± 5.2	13.6 ± 4.2	8.3 ± 3.0	11.7 ± 3.2
α	3.36 ± 0.26	3.57 ± 0.19	3.77 ± 0.26	3.97 ± 0.18	4.15 ± 0.26	4.38 ± 0.18
β_1	-0.17 ± 0.09	-0.13 ± 0.05	-0.23 ± 0.09	-0.17 ± 0.05	-0.29 ± 0.09	-0.22 ± 0.06
β_2 $(\times 10^{-2})$	-0.02 ± 0.50	-0.01 ± 0.50	-0.04 ± 0.43	-0.02 ± 0.43	-0.10 ± 0.46	0.01 ± 0.39
ψ	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01	0.02 ± 0.01
σ	0.65 ± 0.17	0.37 ± 0.09	0.67 ± 0.16	0.36 ± 0.08	0.71 ± 0.17	0.35 ± 0.08

pixels is included in the tumour core. This leads to higher U-values, and hence a larger intercept term in the population model of the core.

The parameters of the tumour annulus population model are more sensitive to the choice of the core–annulus threshold (Table 2). As the threshold increases (e.g. the 97.5th percentile), a larger portion of the tumour with low baseline contrast uptake (compared with other tumour annulus pixels) is excluded from the tumour annulus, which leads to larger U-values (the probability of randomly picking a pixel in the tumour annulus with higher baseline contrast uptake than a healthy pixel is higher), and hence the intercept term in the population model is larger. Furthermore, the change-point location is shifted towards zero, which reflects a steeper decline in the U-values soon after the initiation of radiotherapy.

Since the slopes of the two linear pieces are relatively stable, the temporal profile of the differential tumour–healthy tissue response to radiation is not sensitive to the threshold in the range of threshold values that was investigated, and nor does the main conclusion.

5. Discussion

A notable feature of the longitudinal image model is that it simultaneously models spatial and temporal correlations. In many cases, they are modelled separately or sequentially. In this paper we used hidden MRFs to model the spatial correlation in the structural change of contrast uptake and build the current true contrast uptake on the true value at the previous visit, to model temporal correlation in a manner that is computationally tractable.

This work does not entertain the possibility of a change in size of tumour over time. This study finds that the contrast uptake in tumour core peaks around 3 weeks after the initiation of radiotherapy. It is reasonable to postulate that the physiological changes (e.g. contrast uptake) that are detected within this relatively short period of time precede noticeable changes in tumour size. Therefore, the findings should be robust.

The image model proposed was fitted for each subject separately. We argue that the number of subjects in the motivating study is quite limited compared with the heterogeneity of tumours. Hence, a pixel level population model that adds an additional level to the model hierarchy would not necessarily improve on the current model and would be more demanding computationally. However, we consider a continuous spatial–temporal MRF model (e.g. Descombes *et al.* (1998)) as a legitimate alternative to the model proposed. Both the spatial and the temporal energy function take the form of a Φ-function (Geman and McClure, 1987; Geman and Reynolds, 1992) to avoid oversmoothing of the distinct boundaries in the data. Compared with the model proposed, the model structure is simpler. However, the spatial and temporal parameters in the Φ-function need to be specified separately. In addition the temporal energy function needs to account for the non-evenly spaced scans. These add up to realistic complications to a seemingly simpler model. Computational difficulty is another potential challenge.

Although the longitudinal image model that was discussed in this paper helps to 'denoise' the images by borrowing strength both across neighbouring pixels and over time, the tens of thousands of pixelwise contrast uptake profiles do not provide useful information individually. We summarize the differential tumour–healthy tissue responses with the Mann–Whitney *U*-statistic, which is intuitive and interpretable. The models and computer algorithms that were developed are suitable for either two-dimensional or three-dimensional data. To illustrate the model and algorithm, we analysed the QMRI slice with the largest tumour volume. It is possible that the results could change to a certain degree if the entire brain volume is used.

Extensions of the model proposed include adding covariate information in the longitudinal image model, e.g. radiation dose maps, and specifying more flexible functional forms for the mean model, such as a cubic spline, which better represent the nature of biological response to radiotherapy.

Acknowledgements

Appendix A

We follow Zhang *et al.* (2010) in transdimensional proposals. We first randomly choose between a split and a merge proposal with equal probability, unless M_j is on the boundaries, i.e. M_{min} or M_{max}; then only a split or merge proposal is valid. When a transdimensional move is proposed (with M_j^* components), the new configuration is denoted by $\mathbf{z}_j^* = (z_{1j}^*, \ldots, z_{Nj}^*)^T$ with $z_{ij} \in \{1, 2, \ldots, M_j^*\}$, the new component parameters are $\boldsymbol{\mu}_j^* = (\mu_{1j}^*, \ldots, \mu_{M_j^* j}^*)^T$ and $\boldsymbol{\sigma}_j^* = (\sigma_{1j}^*, \ldots, \sigma_{M_j^* j}^*)^T$. The new true contrast is denoted by $\boldsymbol{\nu}_j^*$.

A.1. Split proposal

If a split proposal is chosen, we randomly pick a component k ($1 \leqslant k \leqslant M_j$) to split, i.e. $P_{select}^{split}(k) = 1/M_j$. We need two extra parameters for the Gaussian mean and variance. To match the increase of dimension, we introduce two independent random variables, $u_1, u_2 \sim \text{beta}(2, 2)$, and define a bijective transformation $(\mu_{k_1}^*, \mu_{k_2}^*, \sigma_{k_1}^{*2}, \sigma_{k_2}^{*2}) = \psi(\mu_k, \sigma_k^2, u_1, u_2)$ that matches the first two moments, $\mu_{k_1}^* = \mu_k - u_1\sigma_k$, $\mu_{k_2}^* = \mu_k + u_1\sigma_k$, $\sigma_{k_1}^{*2} = 2u_2(1 - u_1^2)\sigma_k^2$ and $\sigma_{k_2}^{*2} = 2(1 - u_2)(1 - u_1^2)\sigma_k^2$.

The acceptance rate of a split proposal is the product of prior ratio, likelihood ratio, proposal ratio and the Jacobian of the bijective transformation.

The prior ratio in a split move (from k in M_j components to k_1 and k_2 in $M_j^* = M_j + 1$ components) is

$$\frac{P(M_j^*) P(\beta_j^*) P(\mathbf{Z}_j^*|\beta_j, M_j) \prod_{k=1}^{M_j^*} P(\mu_{kj}^*) P(\sigma_{kj}^{2*})}{P(M_j) P(\beta_j) P(\mathbf{Z}_j|\beta_j, M_j) \prod_{k=1}^{M_j} P(\mu_{kj}) P(\sigma_{kj}^2)} = \frac{\beta_j^{*\,\alpha_\beta - 1} \exp(-\beta_\beta \beta_j^*)}{\beta_j^{\alpha_\beta - 1} \exp(-\beta_\beta \beta_j)} \frac{g^{-1}(\beta_j^*, M_j^*) \exp\left(\sum_{i \sim i'} \beta_j^* I[z_{ij}^* = z_{i'j}^*]\right)}{g^{-1}(\beta_j, M_j) \exp\left(\sum_{i \sim i'} \beta_j I[z_{ij} = z_{i'j}]\right)}$$

$$\times \frac{\sigma_\mu^{-1} \exp\{\sigma_\mu^{-2}(\mu_{k_1 j}^* - \mu_\mu)^2\} \sigma_\mu^{-1} \exp\{-0.5\sigma_\mu^{-2}(\mu_{k_2 j}^* - \mu_\mu)^2\}}{\sigma_\mu^{-1} \exp\{-0.5\sigma_\mu^{-2}(\mu_{kj} - \mu_\mu)^2\}}$$

$$\times \frac{(\sigma_{k_1 j}^{-2*})^{\alpha_\sigma - 1} \exp(-\beta_\beta \sigma_{k_1 j}^{-2*})(\sigma_{k_2 j}^{-2*})^{\alpha_\sigma - 1} \exp(-\beta_\beta \sigma_{k_2 j}^{-2*})}{(\sigma_{kj}^{-2})^{\alpha_\sigma - 1} \exp(-\beta_\beta \sigma_{kj}^{-2})}.$$

The likelihood ratio term in the split proposal is $\Pi_{i=1}^{N} \exp[-0.5\sigma_e^{-2}\{(y_{ij} - \nu_{ij}^*)^2 - (y_{ij} - \nu_{ij})^2\}]$.

We construct the reversible split and merge move on the equivalence classes, i.e. the lattice is partitioned into same-labelled equivalence classes. To compute the proposal ratio, we need to calculate the allocation probabilities that are associated with both moves.

We start with a split move. Let $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_L$ denote the L equivalence classes with label $E_1, E_2, \ldots, E_L$. The allocation probability of class ε_l $(l = 1, 2, \ldots, L)$ assuming label k^* $(k^* = 1, 2, \ldots, M_j + 1)$ in the proposed configuration $\mathbf{z}_j^*$ is

$$p_{k^*}^{l*} = P(Z_{ij}^* = k^*, \text{ for all } i \in \varepsilon_l) \propto \prod_{i \in \varepsilon_l} \exp\{-0.5\sigma_e^{-2}(y_{ij} - \nu_{ij}^*)^2\}.$$

The allocation probability from $\mathbf{z}_j^*$ is $P_{\text{alloc}}(\mathbf{z}_j^*) = \Pi_{l=1}^{L} p_{E_l}^l$. The probability of class ε_l assuming label k $(k = 1, 2, \ldots, M_j)$ satisfies

$$p_k^l = P(Z_{ij} = k, \text{ for all } i \in \varepsilon_l) \propto \prod_{i \in \varepsilon_l} \exp\{-0.5\sigma_e^{-2}(y_{ij} - \nu_{ij})^2\}.$$

The p_k^{l*} satisfy $\Sigma_{k=1}^{M_j+1} p_k^{l*}$ for all $l = 1, 2, \ldots, L$. Let E_l^* $(E_l^* = 1, 2, \ldots, M_j + 1)$ denote the new label of class ε_l drawn on the basis of the above probabilities, i.e. $p_{E_l^*}^{l*} = P(Z_{ij}^* = E_l^*, \forall i \in \varepsilon_l)$. Hence, the allocation probability from configuration $\mathbf{z}_j$ conditional on the equivalence classes is $P_{\text{alloc}}(\mathbf{z}_j) = \Pi_{l=1}^{L} p_{E_l^*}^{l*}$.

The proposal ratio of a split move is

$$\frac{P_{\text{split}}(M_j - 1) P_{\text{select}}^{\text{split}}(k) P_{\text{alloc}}(\mathbf{z}_j^*)}{P_{\text{merge}}(M_j) P_{\text{select}}^{\text{merge}}(k_1, k_2) P_{\text{alloc}}(\mathbf{z}_j)}.$$

A.2. Merge proposal

The merge proposal is determined by the above split move. We randomly pick a pair of components with adjacent means, say k_1 and k_2 with $P_{\text{select}}^{\text{merge}}(k_1, k_2) = 1/(M_j - 1)$. Inverting the bijective transformation in the split proposal, the parameters for the new component are $\mu_k^* = 0.5(\mu_{k_1} + \mu_{k_2})$ and $\mu_k^{*2} + \sigma_k^{*2} = 0.5(\mu_{k_1}^2 + \sigma_{k_1}^2 + \mu_{k_2}^2 + \sigma_{k_2}^2)$. We denote the new set of parameters by $\boldsymbol{\mu}^* = (\mu_1^*, \mu_2^*, \ldots, \mu_{M-1}^*)^{\text{T}}$ and $\boldsymbol{\sigma}^{*2} = (\sigma_1^{*2}, \sigma_2^{*2}, \ldots, \sigma_{M-1}^{*2})^{\text{T}}$. The proposal ratio of a merge move is the reciprocal of the split move.

References

Besag, J. (1974) Spatial interaction and the statistical analysis of lattice systems (with discussion). *J. R. Statist. Soc.* B, **36**, 192–236.

Cao, Y., Tsien, C. I., Shen, Z., Tatro, D. S., Ten Haken, R., Kessler, M. L., Chenevert, T. L. and Lawrence, T. S. (2005) Use of magnetic resonance imaging to assess blood-brain/blood-glioma barrier opening during conformal radiotherapy. *J. Clin. Oncol.*, **23**, 4127–4136.

Chen, H. M., Shao, Q. M. and Ibrahim, J. G. (2000) Monte Carlo methods in Bayesian computation. New York: Springer.

Curran, W. J., Scott, C. B., Horton, J., Nelson, J. S., Weinstein, A. S., Fischbach, A. J., Chang, C. H., Rotman, M., Asbell, S. O. and Krisch, R. E. (1993) Recursive partitioning analysis of prognostic factors in three radiation therapy oncology group malignant glioma trials. *J. Natn. Cancer Inst.*, **85**, 704–710.

Descombes, X., Kruggel, F. and van Cramon, D. Y. (1998) Spatio-temporal fMRI analysis using Markov Random Fields. *IEEE Trans. Med. Imgng*, **17**, 1028–1039.

Gelfand, A. E., Dey, D. K. and Chang, H. (1992) Model determination using predictive distributions with implementation via sampling-based methods. *Bayesn Statist.*, **4**, 147–167.

Gelman, A. and Meng, X. L. (1998) Simulating normalizing constants: from importance sampling to bridge sampling to path sampling. *Statist. Sci.*, **13**, 163–185.

Geman, S. and McClure, D. (1987) Statistical methods for tomographic image reconstruction. *Proc. 46th Sessn Int. Statist. Inst.*, 22–26.

Geman, S. and Reynolds, G. (1992) Constrained restoration and recovery of discontinuities. *IEEE Trans. Pattn Anal. Mach. Intell.*, **14**, 367–383.

Hamstra, D. A., Chenevert, T. L., Moffat, B. A., Johnson, T. D., Meyer, C. R., Mukherji, S. K., Quint, D. J., Gebarski, S. S., Fan, X., Tsien, C. I., Lawrence, T. S., Junck, L., Rehemtulla, A. and Ross, B. D. (2005) Evaluation of the functional diffusion map as an early biomarker of time-to-progression and overall survival in high-grade glioma. *Proc. Natn. Acad. Sci. USA*, **102**, 16759–16764.

Higdon, D. (1998) Auxiliary variable methods for Markov Chain Monte Carlo with applications. *J. Am. Statist. Ass.*, **93**, 585–595.

Lei, T. and Udupa, J. K. (2002) Statistical properties of X-ray CT and MRI from imaging physics to image statistics. *Proc. SPIE Med. Imgng*, **4682**, 82–93.

Liang, Z. P. and Lauterbur, P. C. (1999) *Principles of Magnetic Resonance Imaging: a Signal Processing Perspective*. Bellingham: SPIE.

Moffat, B. A., Chenevert, T. L., Lawrence, T. S., Meyer, C. R., Johnson, T. D., Dong, Q., Tsien, C., Mukherji, S., Quint, D. J., Gebarski, S. S., Robertson, P. L., Junck, L. R., Rehemtulla, A. and Ross, B. D. (2005) Functional diffusion map: a noninvasive MRI biomarker for early stratification of clinical brain tumor response. *Proc. Natn. Acad. Sci. USA*, **102**, 5524–5529.

Ogata, Y. (1989) A Monte Carlo method for high dimensional integration. *Numer. Math.*, **55**, 137–157.

Potts, R. B. (1952) Spontaneous magnetization of a triangular Ising lattice. *Phys. Rev.*, **88**, 352.

Swendsen, R. H. and Wang, J. S. (1987) Nonuniversal critical dynamics in Monte Carlo simulations. *Phys. Rev. Lett.*, **58**, 86–88.

West, M. and Harrison, J. (1999) *Bayesian Forecasting and Dynamic Models*, 2nd edn. New York: Springer.

Zhang, X., Johnson, T. D., Little, R. J. A. and Cao, Y. (2008) Quantitative magnetic resonance image analysis via the EM algorithm with stochastic variation. *Ann. Appl. Statist.*, **2**, 736–755.

Zhang, X., Johnson, T. D., Little, R. J. A. and Cao, Y. (2010) A Bayesian image analysis of radiation induced changes in tumor vascular permeability. *Bayesn Anal.*, **5**, 189–212.