Journal of the Royal Statistical Society
SERIES C Applied Statistics

Appl. Statist. (2012)
61, *Part* 1, *pp.* 83–98

Predicting treatment efficacy via quantitative magnetic resonance imaging: a Bayesian joint model

Jincao Wu, Timothy D. Johnson, Craig J. Galbán, Thomas L. Chenevert, Charles R. Meyer, Alnawaz Rehemtulla, Daniel A. Hamstra and Brian D. Ross

University of Michigan, Ann Arbor, USA

[Received December 2009. Final revision April 2011]

Summary. The prognosis for patients with high grade gliomas is poor, with a median survival of 1 year. Treatment efficacy assessment is typically unavailable until 5–6 months post diagnosis. Investigators hypothesize that quantitative magnetic resonance imaging can assess treatment efficacy 3 weeks after therapy starts, thereby allowing salvage treatments to begin earlier. The purpose of this work is to build a predictive model of treatment efficacy by using quantitative magnetic resonance imaging data and to assess its performance. The outcome is 1-year survival status. We propose a joint, two-stage Bayesian model. In stage I, we smooth the image data with a multivariate spatiotemporal pairwise difference prior. We propose four summary statistics that are functionals of posterior parameters from the first-stage model. In stage II, these statistics enter a generalized non-linear model as predictors of survival status. We use the probit link and a multivariate adaptive regression spline basis. The hybrid Metropolis-within-Gibbs algorithm and reversible jump Markov chain Monte Carlo methods are applied iteratively between the two stages to estimate the posterior distribution. Through both simulation studies and model performance comparisons we find that we can achieve higher overall correct classification rates by accounting for the spatiotemporal correlation in the images and by allowing for a more complex and flexible decision boundary provided by the generalized non-linear model.

Keywords: Bayesian analysis; Image analysis; Multivariate adaptive regression splines; Multivariate pairwise difference prior; Quantitative magnetic resonance imaging; Spatiotemporal model

1. Introduction

Our work is motivated by a need to analyse data collected from quantitative magnetic resonance imaging (QMRI) studies appropriately, and to determine whether QMRI can be used as an early predictor of treatment efficacy as measured by survival for patients with malignant gliomas. The data come from a pilot study of 53 high grade glioma patients (Hamstra *et al.*, 2005). The prognosis for patients with high grade gliomas is poor. The mortality rate, at the time of data collection, is high with a median survival of 1 year after diagnosis (Laws *et al.*, 2003). Treatment is a combined approach of surgery (if possible) and radiation therapy followed by chemotherapy. Assessment of treatment efficacy is based on radiological response approximately 8–10 weeks post therapy, or approximately 5–6 months after diagnosis (Moffat *et al.*, 2005; Hamstra *et al.*, 2008). Radiological response is determined by the change in tumour size from baseline as measured on anatomical MR images. For those with progressive disease,

Address for correspondence: Jincao Wu, Department of Biostatistics, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109–2029, USA.
E-mail: jincaowu@umich.edu

 0035–9254/12/61083

salvage therapy is given. However, it is typically too late for the salvage therapy to have any effect in prolonging survival (Moffat *et al.*, 2005). If treatment efficacy can be assessed earlier, salvage therapies can begin earlier or therapy can be modified.

In the pilot study, two different QMRI studies (diffusion and perfusion) and standard anatomical MRI studies were conducted at each of two time points: baseline (1 week before therapy) and 3 weeks after therapy begins. All four quantitative images were registered to the pretreatment anatomical MR image via a mutual information algorithm (Meyer *et al.* (1997), i.e. an affine translation and rotation). Full imaging data were available on 47 of the 53 patients; therefore we analyse the data from these 47 patients. Tumours were identified on contrast-enhanced T1-weighted MR images at both time points and segmented (outlined) by a radiologist. We use the intersection of the segmented tumours as the region of interest (ROI) (Hamstra *et al.*, 2005, 2008; Moffat *et al.*, 2005). Using the intersection of the segmented tumours, as opposed to the union, avoids the potential comparison of tumour in one image with healthy tissue or oedema in the other image that may occur in the symmetric difference of the segmented tumours because of small changes in tumour volume, swelling of tissue caused by therapy and errors in segmentation.

The apparent diffusion coefficient (ADC) is a measure of the magnitude of Brownian motion of water molecules in the extracellular space of tissue (Hamstra *et al.*, 2005, 2008; Moffat *et al.*, 2005). Diffusion in biological systems is a complex phenomenon, influenced directly by tissue microstructure. Its measurement can provide information about the organization of this structure in normal and diseased tissue (Basser and Jones, 2002). As tumour cells lyse, the ratio of extracellular to intracellular fluid increases, thus causing a temporary increase in ADC (Moffat *et al.*, 2005, 2006). Perfusion is a measure of tissue-specific blood flow and blood volume and reflects the delivery of essential nutrients to tissue (Galbán *et al.*, 2009). It is hypothesized that effective therapy will disrupt tumour blood supply by damaging tumour neovascularity, resulting in decreased tumour perfusion. Furthermore, recent studies have suggested that QMRI can be used for early prediction of therapeutic efficacy. Early changes detected in mean tumour ADC values were first found to be correlated with treatment response in rodent tumour models (Ross *et al.*, 1994; Zhao *et al.*, 1996; Chinnaiyan *et al.*, 2000). Previous studies investigating perfusion MRI for tumour diagnosis and response monitoring relied on the whole-tumour mean value as the summary statistic of the perfusion maps for quantification of haemodynamic parameters, with varying success (Young *et al.*, 2007; Law *et al.*, 2007). The functional diffusion map (FDM), a voxel-by-voxel approach, was recently reported as an early quantitative biomarker for clinical brain tumour treatment outcome (Hamstra *et al.*, 2005, 2008; Moffat *et al.*, 2005). Galbán *et al.* (2009) have also shown that the functional perfusion map based on perfusion MRI (obtained in the same way as the FDM) is predictive of overall survival. However, both the FDM and the functional perfusion map treat voxels as independent observations, thus ignoring spatial structure in the images. Treating the data as independent observations may lead to incorrect variance estimates and invalid inference. Our work is motivated by all of these studies and aims to build a statistically robust and predictive model for treatment efficacy based on both the ADC and the relative cerebral blood flow (RCBF) (which is a measure of perfusion; Galbán *et al.* (2009)) images. An axial slice of a registered ADC image, an RCBF and a T1-weighted, contrast-enhanced MR image are shown in Fig. 1.

We propose a joint, two-stage Bayesian predictive model. In the first stage, we smooth the images (two images at each of two time points) by using a multivariate pairwise difference prior that models the spatiotemporal correlation in the images. The pairwise difference prior was first introduced by Besag (1993). It is a member of the class of pairwise interaction Markov random-field models and captures general and local characteristics of the image. *A priori*, it

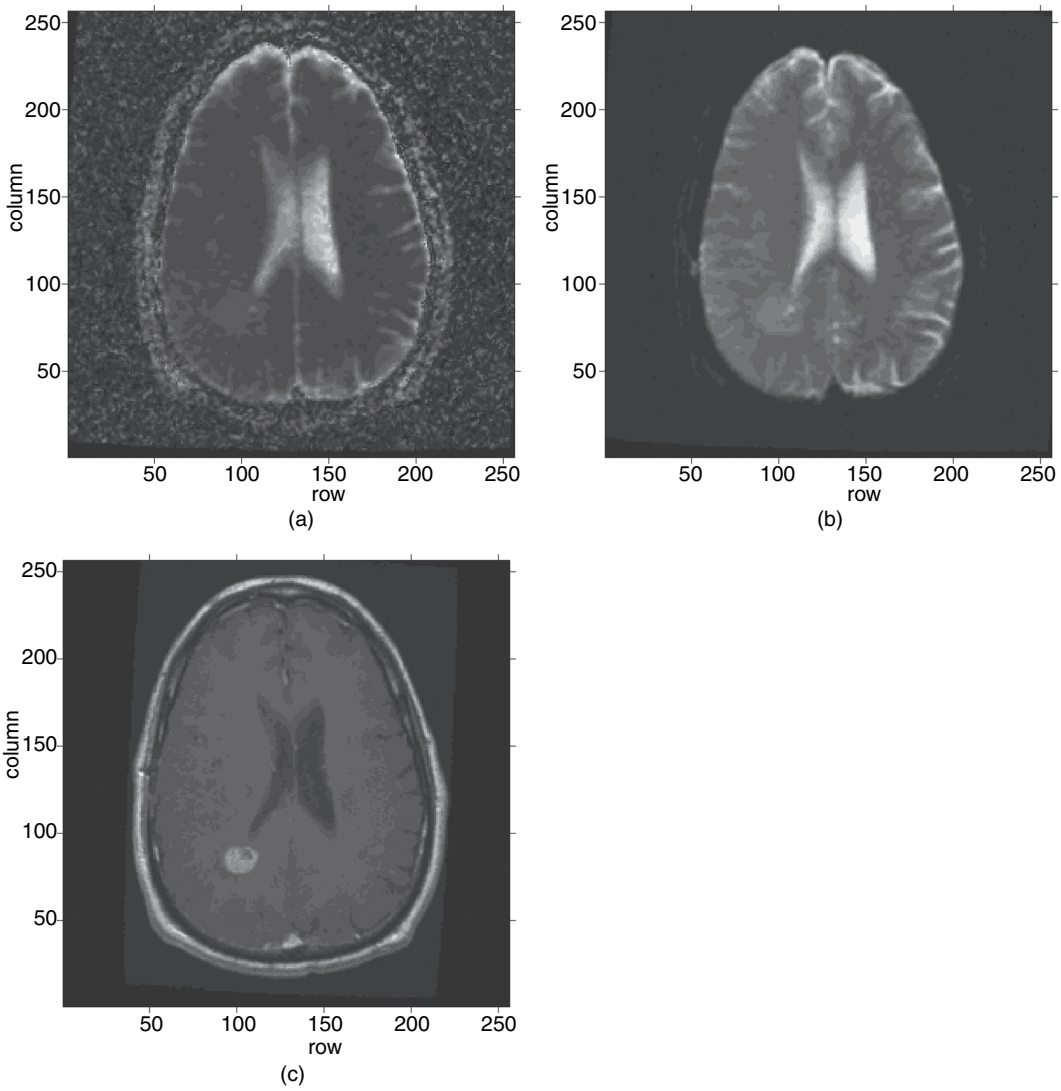

Fig. 1. Single axial slice of pretreatment MRI data (the tumour is visible in all three images; it is located at roughly voxel (100,80) just below the left ventricle): (a) diffusion MR image; (b) perfusion MR image; (c) T1-weighted contrast-enhanced MR image

assumes that the mean values of neighbouring voxels are positively correlated. We extend the pairwise difference prior to the multivariate setting. We then propose four summary statistics that are functionals of the parameters in stage I. The statistics enter the second-stage model as predictors of 1-year survival status. The second-stage model is a generalized non-linear model (GNLM) that was proposed by Holmes and Denison (2003). The GNLM uses a probit link, for computational efficiency, and a Bayesian multivariate adaptive regression spline (BMARS) basis. The MARS model was introduced by Friedman (1991). The BMARS basis allows the predictors to enter the GNLM model non-linearly, thus allowing for a very flexible decision boundary. The two models are fitted jointly and the model is validated via cross-validated prediction. Algorithmically, the models are joined by iterating between the two stages in a generalized

Markov chain Monte Carlo (MCMC) simulation (Metropolis-within-Gibbs updates in stage I and a hybrid reversible jump MCMC and Gibbs updates of hyperparameters in stage II).

Compared with current methods, our joint model has several new features and improvements. In the first stage, our model

(a) accounts for spatiotemporal correlation in the images, as well as the correlation between the ADC and RCBF images,
(b) increases the signal-to-noise ratio by smoothing the images and
(c) reduces the data dimension via subject level summary statistics.

In the second stage, our model allows for a more flexible classification boundary than that allowed by the standard linear systematic component of a generalized linear model. The joint model that we propose propagates the sampling error from stage I into stage II. We adopt the Bayesian paradigm for estimating and predicting outcomes. Furthermore, model uncertainty is captured by model averaging.

This paper is organized as follows. In Section 2, we first outline, at a high level, our joint model and then specify the two stages of the model and propose our model evaluation strategy that we have adopted. The pilot study data are then analysed in Section 3. We show that our model outperforms simpler models in Section 4. The paper concludes with a discussion, summarizing the strengths and limitations of our approach. For brevity, results from simulation studies and sensitivity analyses can be found in a supplementary Web-based material (SWBM) document that is available at `http://www.bepress.com/umichbiostat/paper83` or by contacting the authors. Detailed mathematical derivations of the posterior distributions, details of the algorithm and pseudocode are also provided in the SWBM.

2. Bayesian joint model

To begin, we briefly describe the joint model. Let $\mathcal{Y}$ denote the set of all images for all subjects and let $\mathbf{Z}$ denote the 1-year survival status (1, dead; 0, alive) vector. Let $\boldsymbol{\Omega} = \boldsymbol{\Omega}_1 \cup \boldsymbol{\Omega}_2$ denote the set of all model parameters where $\boldsymbol{\Omega}_1$ is the set of stage I model parameters and predictive values and $\boldsymbol{\Omega}_2$ is the set of stage II model parameters. We further note that the set (over all subjects) of all summary statistics, $\mathcal{X}$, calculated in stage I is a functional vector of $\boldsymbol{\Omega}_1$ and that $\mathbf{Z}$ depends on $\boldsymbol{\Omega}_1$ only through $\mathcal{X} = F(\boldsymbol{\Omega}_1)$. The posterior distribution can be factored as

$$\pi(\boldsymbol{\Omega}|\mathcal{Y}, \mathbf{Z}) \propto \pi(\mathcal{Y}|\boldsymbol{\Omega}_1)\, \pi(\boldsymbol{\Omega}_1)\, \pi(\mathbf{Z}|\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)\, \pi(\boldsymbol{\Omega}_2). \tag{1}$$

We shall use $\pi(\mathbf{Z}|\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)$, $\pi(\mathbf{Z}|\boldsymbol{\Omega})$ and $\pi(\mathbf{Z}|\mathcal{X}, \boldsymbol{\Omega}_2)$ interchangeably depending on the context. We draw from the posterior (1) via MCMC simulation by iteratively drawing between the full conditional distribution of $\boldsymbol{\Omega}_1$,

$$\pi(\boldsymbol{\Omega}_1|\mathcal{Y}, \mathbf{Z}, \boldsymbol{\Omega}_2) \propto \pi(\mathcal{Y}|\boldsymbol{\Omega}_1)\, \pi(\boldsymbol{\Omega}_1)\, \pi(\mathbf{Z}|\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2), \tag{2}$$

and the full conditional distribution of $\boldsymbol{\Omega}_2$,

$$\pi(\boldsymbol{\Omega}_2|\mathbf{Z}, \boldsymbol{\Omega}_1) \propto \pi(\mathbf{Z}|\boldsymbol{\Omega}_1, \boldsymbol{\Omega}_2)\, \pi(\boldsymbol{\Omega}_2). \tag{3}$$

The full conditionals in expressions (2) and (3) are easily derived from expression (1) and repeated use of Bayes's theorem. We assume, *a priori*, that $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are independent. However, they are dependent in the posterior. From equations (2) and (3), the full conditional distribution of $\boldsymbol{\Omega}_1$ depends on $\mathbf{Z}$, $\mathcal{Y}$ and $\boldsymbol{\Omega}_2$ and the full conditional distribution of $\boldsymbol{\Omega}_2$ is conditionally independent of $\mathcal{Y}$ given $\boldsymbol{\Omega}_1$.

The remainder of this section is broken up into four subsections. In Section 2.1 we define the multivariate pairwise difference prior model, in Section 2.2 we define the GNLM and then in Section 2.3 we give an overview of how we sample from the joint posterior distribution that is specified in expression (1). In the last subsection, Section 2.4, we describe how we evaluate our model.

2.1. Stage I

In this subsection, patients' subscripts are suppressed to reduce the notational burden. Tumour voxels (i.e. volume elements—a cube) are indexed by $i = 1, 2, \ldots, n$, where the tumour size n (i.e. the size of the tumour ROI defined in Section 1) ranges from 770 to 20380 voxels with a mean of 6143 and standard deviation of 4721. Two voxels, i and i', that share a common face are called neighbours, denoted by $i \sim i'$. Let $N_i = \{i' : i' \sim i\}$ denote the set of neighbours of voxel i with $|N_i|$ denoting the number of neighbours. Let Y_{ith} represent the image intensity at voxel i, time $t = 1, 2$ (baseline and week 3 respectively) and image type h (h =1, diffusion; $h = 2$, perfusion). The vector of image intensities at voxel i is $\mathbf{Y}_i = (Y_{i11}, Y_{i12}, Y_{i21}, Y_{i22})^{\mathrm{T}}$. We split $\mathbf{Y}_i$ into two subvectors by time: $\mathbf{Y}_{it} = (Y_{it1}, Y_{it2})^{\mathrm{T}}$. Furthermore, let $\mathbf{Y} = (\mathbf{Y}_1^{\mathrm{T}}, \ldots, \mathbf{Y}_n^{\mathrm{T}})^{\mathrm{T}}$. Each Y_{ith} is measured with error with mean μ_{ith}. Let $\boldsymbol{\mu}_i = (\mu_{i11}, \mu_{i12}, \mu_{i21}, \mu_{i22})^{\mathrm{T}}$ with corresponding subvectors $\boldsymbol{\mu}_{it} = (\mu_{it1}, \mu_{it2})^{\mathrm{T}}$. Let $\boldsymbol{\mu} = (\boldsymbol{\mu}_1^{\mathrm{T}}, \ldots, \boldsymbol{\mu}_n^{\mathrm{T}})^{\mathrm{T}}$. Note that the components in $\mathbf{Y}_i$ are correlated with covariance Σ.

We extend Besag's (1993) pairwise difference prior model to the multivariate setting. First,

$$[\mathbf{Y}|\boldsymbol{\mu}, \Sigma^*] \sim N(\boldsymbol{\mu}, \Sigma^*), \tag{4}$$

where $\Sigma^* = \mathrm{diag}(\Sigma)$—a block diagonal matrix with Σ along the main diagonal. The prior distribution of the mean vector $\boldsymbol{\mu}$ is

$$\pi(\boldsymbol{\mu}|\Psi) \propto \exp\Big\{ -0.5 \sum_{i \sim i'} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i'})^{\mathrm{T}} \Psi^{-1} (\boldsymbol{\mu}_i - \boldsymbol{\mu}_{i'}) \Big\}.$$

Spatial correlation in the image is modelled through the diagonal elements of the 4×4 covariance matrix Ψ. The off-diagonal elements of Ψ account for temporal correlation within an image type, correlation between image types at a particular time and correlation over time and across image types. The covariance matrix Σ accounts for residual covariances. In equation (4), we assume the normal distribution for the noise in the images, which is commonly used and has been justified in Lei and Udupa (2002) and Liang *et al.* (2000).

A priori, Σ and Ψ are assigned inverse Wishart distributions: $W^{-1}(I_4, 5)$. The scale matrix I_4 is the 4×4 identity matrix and the number of degrees of freedom is 5. The degrees of freedom can be regarded as the *a priori* sample size. Given the large n, this results in a rather weak prior.

2.1.1. Predicting tumour response under the 'null'

Ideally we would compare the observed tumour response with its counterfactual: tumour response given no treatment. Given that this is impossible, our summary statistics will be based on comparing the observed tumour response with the predicted tumour response in the contralateral hemisphere of the brain under the assumption that the change in ADC or RCBF values in healthy tissue in the contralateral brain and those of tumour in the contralateral brain, if they could be observed, are similar. In the contralateral brain the healthy tissue receives a low dose of radiation and little damage from chemotherapy because of the blood–brain barrier which blocks large chemotherapy molecules. Thus, the healthy tissue in the contralateral brain is pro-

tected from therapy and diffusion and perfusion are stable over the short time period between imaging sessions. We define a healthy tissue ROI in the contralateral brain. The healthy tissue ROI is obtained by reflecting the tumour ROI, approximately about the midline of the brain, to the contralateral hemisphere. We then ensure, visually, that the healthy tissue ROI lies within the grey matter of the brain (some white matter is fine). If the healthy tissue ROI intersects the ventricles, meninges or skull, we manually shift the ROI, to avoid this overlap (details can be found in the SWBM). We now describe how we predict tumour response in the healthy tissue ROI, which we refer to as the *null response*.

First we build a multivariate pairwise difference prior model for the healthy tissue data in the healthy tissue ROI. The model is identical to that described above with the following notational changes. For healthy tissue, in the healthy tissue ROI, let $\mathbf{W}_i$ denote the image intensities for voxel i with mean vector $\boldsymbol{\nu}_i$. The covariance of the $\mathbf{W}_i$ will be denoted Δ and the covariance of the mean vector $\boldsymbol{\nu}_i$ will be denoted Ω. The number of voxels in the healthy tissue ROI is also n. Denote the set of voxels in the healthy ROI by $\mathcal{H}$. We extend the healthy tissue ROI by a one-voxel-thick shell and denote the set of voxels in this shell by $\mathcal{S}$. Without this extension, $\tilde{\mathbf{Y}}_{i2}$ and $\tilde{\boldsymbol{\mu}}_{i2}$ are not identifiable (see equations (5) and (6) below and section 5 in the SWBM). Let n^{s} denote the number of voxels in the shell and let $n^{\mathrm{e}} = n + n^{\mathrm{s}}$ be the number in the extended ROI. Let $N_i^{\mathrm{e}} = \{i' : i' \sim i\}$ denote the set of neighbours of voxel i in the extended ROI and $|N_i^{\mathrm{e}}|$ denote the number in this set.

Now, to predict the tumour null response, we translate the tumour baseline values $\mathbf{Y}_{i1}$ to the healthy tissue ROI, using the same reflection and shift that created the healthy tissue ROI. We partition the 4×4 covariance matrices into four 2×2 matrices. The multivariate pairwise difference prior for prediction is

$$\left(\begin{pmatrix} \mathbf{Y}_{i1} \\ \tilde{\mathbf{Y}}_{i2} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \tilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \right) \sim N\left\{ \begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \tilde{\boldsymbol{\mu}}_{i2} \end{pmatrix}, \begin{pmatrix} \Delta_{11} & \Delta_{12} \\ \Delta_{21} & \Delta_{22} \end{pmatrix} \right\} \tag{5}$$

where $\tilde{\mathbf{Y}}_{i2} = (\tilde{Y}_{i21}, \tilde{Y}_{i22})^{\mathrm{T}}$ is the predicted null response at time point 2 and $\tilde{\boldsymbol{\mu}}_{i2}$ is its mean. The $\boldsymbol{\mu}_{i1}$ are obtained from the posterior distribution of the tumour multivariate pairwise difference prior model and the covariances from the healthy tissue multivariate pairwise difference prior models. Let

$$\begin{pmatrix} \boldsymbol{\mu}_{i1}^{*} \\ \tilde{\boldsymbol{\mu}}_{i2}^{*} \end{pmatrix} = |N_i^{\mathrm{e}}|^{-1} \left\{ \sum_{i' \in N_i^{\mathrm{e}} \cap \mathcal{S}} \begin{pmatrix} \mathbf{0} \\ \boldsymbol{\nu}_{i'2} \end{pmatrix} + \sum_{i' \in N_i^{\mathrm{e}} \cap \mathcal{H}} \begin{pmatrix} \boldsymbol{\mu}_{i'1} \\ \tilde{\boldsymbol{\mu}}_{i'2} \end{pmatrix} \right\}. \tag{6}$$

The prior for the mean vector in expression (5) is

$$\left(\begin{pmatrix} \boldsymbol{\mu}_{i1} \\ \tilde{\boldsymbol{\mu}}_{i2} \end{pmatrix} \middle| \begin{pmatrix} \boldsymbol{\mu}_{i1}^{*} \\ \tilde{\boldsymbol{\mu}}_{i2}^{*} \end{pmatrix}, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{i22} \end{pmatrix} \right) \sim N\left\{ \begin{pmatrix} \boldsymbol{\mu}_{i1}^{*} \\ \tilde{\boldsymbol{\mu}}_{i2}^{*} \end{pmatrix}, |N_i^{\mathrm{e}}|^{-1} \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \right\} \tag{7}$$

where the covariances are obtained from the posterior of the healthy tissue multivariate pairwise difference prior model. The covariances taken from the posterior of the healthy tissue multivariate pairwise difference prior model describe the spatiotemporal relationship between the baseline tumour ADC or RCBF values and the predicted values under our assumption that tumour changes would be similar to healthy tissue changes in the environment of the contralateral hemisphere. We need to ensure that Σ_{11} and Δ_{11} are similar as well as Ψ_{11} and Ω_{11} as these describe the baseline residual covariances and spatial covariances. If they are much different, the inequality in the baseline covariances may result in biased predictions. We may be tempted to replace Δ_{11} with Σ_{11} in expressions (5) and Ω_{11} with Ψ_{11} in expression (7); however, there is no guarantee that the resulting covariance matrices would be positive definite. After

fitting our model to the data we investigated whether these assumptions hold by comparing the posterior expected values of these leading submatrices. To compare them, we computed the relative root-mean-squared difference between the three unique elements in the leading 2×2 submatrices, where the mean is computed over draws from the posterior (see details in the SWBM, section 1.1). The relative root-mean-squared difference between the leading 2×2 submatrices of Δ and Σ (relative to Δ) is 0.038 (standard deviation 0.029) and that between the leading submatrices of Ω and Ψ (relative to Ω) is 0.039 (standard deviation 0.018)—both small relative differences—hence we feel that this assumption is justified in our model.

Now we explicitly define the stage I parameter set $\boldsymbol{\Omega}_1$ and, at the same time, add a subject-specific index, j. Gather all parameters and predictive values into a set of parameters for subject j: $\boldsymbol{\Omega}_{1j} = \{\{\boldsymbol{\mu}_{i,j}\}_{i=1}^{n_j}, \{\boldsymbol{\nu}_{i,j}\}_{i=1}^{n_j}, \Sigma_j, \Psi_j, \Omega_j, \Delta_j, \{\tilde{\mathbf{Y}}_{i2,j}\}_{i=1}^{n_j}, \{\tilde{\boldsymbol{\mu}}_{i2_j}\}_{i=1}^{n_j}\}$. Then $\boldsymbol{\Omega}_1 = \cup_j \boldsymbol{\Omega}_{1j}$.

2.1.2. Summary statistics

The summary statistics are based on comparing the observed tumour response with the predicted tumour response under the null hypothesis. Previous work suggests that the mean change in tumour ADC values is not predictive of treatment efficacy in humans (Chenevert *et al.*, 2000; Moffat *et al.*, 2005). Empirically, however, the baseline tumour ADC (RCBF) histogram and the week 3 tumour ADC (RCBF) histogram are notably different. This gave us the idea to investigate whether the Kullback–Leibler divergence (Kullback and Leibler, 1951) between the posterior and predictive draws of μ_{i2h} and $\tilde{\mu}_{i2h}, h = 1, 2$ respectively, would be good predictors of treatment efficacy. Specifically, we draw μ_{i2h} from its full conditional posterior for all i in the tumour ROI and create a histogram and draw $\tilde{\mu}_{i2h}$ from its full conditional posterior for all i, create a histogram and then compute the Kullback–Leibler divergence between these two histograms (see details in the SWBM, section 1.2). We denote the statistics as dKLD for the diffusion image and pKLD for the perfusion image.

Hamstra *et al.* (2005, 2008) and Moffat *et al.*, 2005) have demonstrated that the FDM, which is a statistical approach for segmenting tumours into regions of response and non-response, on the basis of a defined upper threshold of ADC change following therapy, is a good biomarker for predicting early tumour response to therapy (this threshold is basically an upper prediction limit of the regression slope of the week 3 tumour ADC values regressed on the baseline ADC values). The FDM approach is based on the rationale that early ADC changes due to therapy are heterogeneous within the tumour. Parts of the tumour respond to therapy and show an increase in ADC, whereas other regions show no change or even a decrease in ADC. However, successful therapy should result in tumour cells lysing with a corresponding increase in ADC: thus the rationale for defining an upper threshold. Furthermore, a successful treatment should result in a decrease in RCBF, as discussed in Section 1. However, again, tumour response is heterogeneous and the mean change is minimal, whereas changes in the tails of the distribution are more pronounced. Inspired by the FDM, we sought statistics that summarize the proportion of extreme expected values, $\boldsymbol{\mu}_{i2}$, in the tumour response relative to the conditional distribution (SWBM, formula (29)) of means of predicted null tumour voxel responses. We propose two additional summary statistics: the conditional diffusion statistic cDS and the conditional perfusion statistic cPS. The first, cDS, is defined as the proportion of tumour voxels that have a mean response that is greater than the 0.975-quantile of the conditional distribution (SWBM, formula (29)) of the same voxel under the null assumption:

$$\text{cDS} = n^{-1} \sum_{i=1}^{n} I[\mu_{i21} > q_{0.975}(\tilde{\mu}_{i21})],$$

where $I[\cdot]$ is the indicator function and $q_{0.975}(\tilde{\mu}_{i21})$ is the 0.975-quantile of the conditional

posterior distribution of $\tilde{\mu}_{i21}$. The summary measure cPS is similarly defined:

$$\text{cPS} = n^{-1} \sum_{i=1}^{n} I[\mu_{i22} < q_{0.025}(\tilde{\mu}_{i22})],$$

where $q_{0.025}(\tilde{\mu}_{i22})$ is the 0.025-quantile of the conditional posterior distribution of $\tilde{\mu}_{i22}$.

2.2. Stage II

For stage II, we borrow the GNLM with a BMARS basis (GNLM–BMARS model) that was proposed by Holmes and Denison (2003) to predict patients' 1-year survival status. For patient j, let $\mathbf{X}_j = (X_{j1}, \ldots, X_{j4})^{\mathrm{T}}$ denote the vector of the summary statistics that is obtained in stage I. Hence $\mathcal{X} = \cup_j \{\mathbf{X}_j\}_j$. Let Z_j index the survival status of patient j, with $Z_j = 1$ representing the death of patient j within 1 year, and $Z_j = 0$ otherwise, for $j = 1, \ldots, M$. Set $\mathbf{Z} = (Z_1, \ldots, Z_M)$. The set of all GNLM–BMARS parameters, $\boldsymbol{\Omega}_2$, will now be subscripted by K, the number of BMARS bases, as the number of bases is treated as a parameter to be estimated and the number of parameters in $\boldsymbol{\Omega}_{2K}$ depends on K. All parameters in $\boldsymbol{\Omega}_{2K}$ will be defined shortly. The GNLM–BMARS model with K basis functions is

$$\begin{aligned} \pi(Z_j = 1 | \mathbf{X}_j, \boldsymbol{\Omega}_{2K}) &= g(\eta_{jK}), \qquad & \eta_{jK} &= \sum_{k=0}^{K} \beta_k B_k(\mathbf{X}_j), \\ B_k(\mathbf{X}_j) &= \begin{cases} 1, & k = 0, \\ \prod_{l=1}^{L_k} [s_{lk}(X_{jw_{lk}} - t_{lk})]_+, & k = 1, 2, \ldots, K. \end{cases} \end{aligned} \tag{8}$$

The link function g could be the cumulative distribution function from any of the commonly used distributions for modelling binary data such as the logistic, normal or extreme value distributions. Owing to the flexibility in the decision boundary that is afforded by the BMARS basis, we argue that the choice of link function is not crucial. Thus, for computational efficiency and simplicity, we use the probit link function, $g(\cdot) = \Phi(\cdot)$, where Φ is the standard normal cumulative distribution function. The function $[\cdot]_+ = \max(0, \cdot)$. K is the number of basis functions in the model. L_k is the degree of interaction in basis function $B_k(\cdot)$. For our application, we set the highest order of interaction to 2. Thus, only main effects and two-way interactions are allowed to enter the model. Estimating higher order interactions with any certainty would require a larger amount of data owing to the curse of dimensionality (Denison *et al.*, 2002). The variable s_{lk} is a sign indicator, taking values in $\{-1, 1\}$; t_{lk} is the location of the spline knot that is associated with the covariate indexed by $w_{lk} \in \{1, 2, 3, 4\}$. Further t_{lk} is restricted to the set of covariate values $\{X_{1w_{lk}}, \ldots, X_{Mw_{lk}}\}$ and all w_{lk} are distinct for each k (i.e. each basis function is at most linear in any one variable). Consult Holmes and Denison (2003) and Denison *et al.* (2002), chapter 4, or the SWBM for further details. Let $\boldsymbol{\beta}_K = (\beta_0, \ldots, \beta_K)^{\mathrm{T}}$ where β_0 is the model intercept. Also, let $\mathbf{L}_K = \{L_1, \ldots, L_K\}$, $\mathbf{s}_K = \{s_{11}, \ldots, s_{L_K K}\}$, $\mathbf{w}_K = \{w_{11}, \ldots, w_{L_K K}\}$, $\mathbf{t}_K = \{t_{11}, \ldots, t_{L_K K}\}$ and $\Theta_K = \{K, \mathbf{s}_K, \mathbf{w}_K, \mathbf{t}_K, \mathbf{L}_K\}$. Then $\boldsymbol{\Omega}_{2K} = \Theta_K \cup \{\boldsymbol{\beta}_K\}$.

We specify non-informative prior distributions for all parameters,

$$\begin{gathered} \pi(L_k = 1) = \pi(L_k = 2) = \tfrac{1}{2}, \\ \pi(w_{1k} = w | L_K = 1) = \tfrac{1}{4}, \qquad w = 1, 2, 3, 4, \\ \pi[(w_{1k}, w_{2k}) = (w, w') | L_K = 2] = \tfrac{1}{6}, \qquad (w, w') = (1,2), (1,3), (1,4), (2,3), (2,4), (3,4), \\ \pi(t_{lk} = X_{jw_{lk}} | w_{lk}) = 1/M, \qquad j = 1, \ldots, M, \end{gathered}$$

$$\pi(s_{lk}=-1)=\pi(s_{lk}=1)=\tfrac{1}{2},$$
$$[\boldsymbol{\beta}_K|v,K]\sim N(0,vI_{K+1}),\qquad [v^{-1}]\sim\text{gamma}(0.001,0.001),$$

with one exception: $[K|\lambda]\sim\text{Poisson}(\lambda)$ and $[\lambda]\sim\text{gamma}(1,0.2)$. We assess the effect of this prior on classification results in the SWBM.

2.3. Sampling from the joint posterior

Now we outline how we sample from the joint posterior given in expression (1). For more details and derivations, consult the SWBM.

We begin with the sampling of stage I parameters. We sample parameters $\boldsymbol{\mu}_{i,j}$, $\boldsymbol{\nu}_{i,j}$, $\tilde{\boldsymbol{\mu}}_{i2,j}$, Σ_j, Ψ_j, Δ_j and Ω_j for $i=1,\ldots,n_j$ and $j=1,\ldots,M$ from their full conditional distributions via a hybrid Metropolis-within-Gibbs algorithm. Both Ψ_j and Σ_j are drawn directly from their full conditionals (inverse Wishart distributions). The remaining parameters are drawn from their full conditionals via Metropolis–Hastings updates (Hastings, 1970). Full details are provided in the SWBM, section 1. We note here that all parameters in a stand-alone multivariate pairwise difference prior model can be updated by a Gibbs algorithm. However, owing to the joint nature of our full model, all stage 1 parameters other than Ψ_j and Σ_j are linked to stage II through the summary statistics and thus require Metropolis-within-Gibbs updates as the full conditionals no longer have a nice distributional form.

Now we outline our posterior sampling algorithm for stage II parameters. In probit regression models, the posterior distribution can be simulated by a Metropolis–Hastings algorithm. However, to simplify computation, Albert and Chib (1993) derived a data augmentation scheme which relies on the latent variable model representation of a binary variable. This approach greatly simplifies sampling from the posterior distribution as the model is transformed from a probit regression model into an equivalent linear model; thus the parameter vector $\boldsymbol{\beta}_K$ can be drawn from its full conditional as opposed to the Metropolis–Hastings algorithm.

We introduce a continuous latent vector $\mathbf{d}=(d_1,\ldots,d_M)^{\mathrm{T}}$. Define the conditional distribution of Z_j given d_j by

$$\pi(Z_j=1|d_j)=\begin{cases}1 & \text{if } d_j>0,\\ 0 & \text{if } d_j\leqslant 0.\end{cases}\tag{9}$$

The full conditional distribution of d_j is straightforward to derive (see the SWBM and Holmes and Denison (2003)) and is

$$[d_j|Z_j=z_j,\mathbf{X}_j,\boldsymbol{\Omega}_{2K}]\sim\begin{cases}N(\eta_{jK},1)\text{ truncated at the left by 0 if } z_j=1,\\ N(\eta_{jK},1)\text{ truncated at the right by 0 if } z_j=0.\end{cases}\tag{10}$$

We draw d_j, $j=1,\ldots,M$, from distribution (10). We then draw $\boldsymbol{\beta}_K$ from its full conditional distribution: $[\boldsymbol{\beta}_K|\mathbf{d},v,\Theta_K,\mathcal{X}]\sim N(\mathbf{m}_K^*,V_K^*)$, where $V_K^*=\{(vI_{K+1})^{-1}+B_K^{\mathrm{T}}B_K\}^{-1}$ and $\mathbf{m}_K^*=V_K^*B_K^{\mathrm{T}}\mathbf{d}$. Standard conjugacy results state that the full conditional distribution of v^{-1} is $[v^{-1}|\boldsymbol{\beta}_K,K]\sim$ gamma$\{0.001+0.5(K+1),0.001+0.5\boldsymbol{\beta}_K^{\mathrm{T}}\boldsymbol{\beta}_K\}$.

All parameters contained in Θ_K are updated via the reversible jump MCMC algorithm (Green, 1995). Since K is random, the dimension of Θ_K varies as well as the column dimension of the matrix of BMARS bases, B_K, and the dimension of the vector $\boldsymbol{\beta}_K$. At each iteration of the algorithm, we randomly (with equal probability) choose to add a new basis function (birth step) or to remove one of the existing basis functions (death step). Thus, covariates (summary statistics) and any two-way interactions enter the model via these birth and death steps. Details of the

reversible jump MCMC algorithm and pseudocode for sampling from the posterior distribution of our joint model are given in the SWBM.

2.4. Model evaluation

The traditional way to evaluate classification models is by randomly partitioning the data into a training set for model building and a test set for model evaluation. However, because of the small sample size in our data set, we evaluate our proposed joint model via cross-validation. To implement cross-validation, a straightforward, but computationally expensive, approach is to run the algorithm multiple times with one observation left out each time. Instead, we adopt the importance sampling approach that was proposed by Gelfand *et al.* (1992) whereby one need only to estimate the posterior distribution of the parameters given the full data set and then by importance sampling compute the predictive probability that $Z_j = 1$ given $\mathbf{Z}_{\{-j\}}$ and $\mathcal{Y}$ for subject j where $\mathbf{Z}_{\{-j\}}$ denotes all observations except that of subject j. Let $\boldsymbol{\Omega}_2^{(t)}$ denote the value of $\boldsymbol{\Omega}_{2K}$ from the tth draw from the posterior and that of $\boldsymbol{\Omega}_1$ by $\boldsymbol{\Omega}_1^{(t)}$. The cross-validated posterior predictive probability is estimated by MCMC output (see the SWBM) and is given by

$$\pi(Z_j = 1|\mathbf{Z}_{\{-j\}}, \mathcal{Y}) = \frac{\sum_{t=1}^{T} \pi(Z_j = 1|\boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)}) / \pi(Z_j = z_j|\boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)})}{\sum_{t=1}^{T} 1/\pi(Z_j = z_j|\boldsymbol{\Omega}_{1j}^{(t)}, \boldsymbol{\Omega}_2^{(t)})}, \tag{11}$$

where z_j is the observed value of Z_j. We assume that the losses that are incurred by a false negative and a false positive prediction are equal. Thus, if $\pi(Z_j = 1|\mathbf{Z}_{\{-j\}}, \mathcal{Y}) \geqslant 0.5$, then the cross-validated prediction of Z_j equals 1 and 0 otherwise.

Although not part of model evaluation, here is a good place to discuss the predictive decision boundary. Theoretically, we could use the predictive distribution, $\pi(Z_{\text{new}} = 1|\mathcal{Y}_{\text{new}}, \mathbf{Z}, \mathcal{Y})$, to define the decision boundary by varying $\mathcal{Y}_{\text{new}}$ over the space of all images. Obviously this is too daunting a task. Instead we shall define the conditional predictive decision boundary in terms of the summary statistics. This decision boundary is a hypersurface in $\mathbb{R}^4$—the covariate space. It is defined as all solutions, $\mathbf{X}_{\text{new}}$, to the posterior predictive probability

$$\tfrac{1}{2} = \pi(Z_{\text{new}} = 1|\mathbf{Z}, \mathbf{X}_{\text{new}}) = \int \pi(Z_{\text{new}} = 1|\mathbf{X}_{\text{new}}, \boldsymbol{\Omega}_{2K})\, \pi(\boldsymbol{\Omega}_{2K}|\mathbf{Z}, \boldsymbol{\Omega}_1)\, \mathrm{d}\boldsymbol{\Omega}_{2K}.$$

We cannot visualize this decision boundary either as the dimension is 4. Therefore, to visualize the decision boundary, we shall marginalize over pairs of covariates and plot the marginal predictive probability map as a function of the remaining pair of covariates by discretizing the marginal covariate space into a grid of values. The marginal decision boundary, then, is a curve in two-dimensional space (Fig. 2).

We note here that at each iteration the number of BMARS bases may change; thus implicit in the estimation of the cross-validated predictive probability and in building the marginal probability maps we average over all potential BMARS models. By doing so, we account for model uncertainty in our results along with the uncertainty in the model parameters and thus, inductively, the uncertainty in the covariates $\mathcal{X}$ (Raftery *et al.*, 1996).

3. Results

Stage I is computationally much more expensive than stage II owing to the large number of voxels, n, in each patient's tumour. We run the algorithm (stage I and stage II combined) for

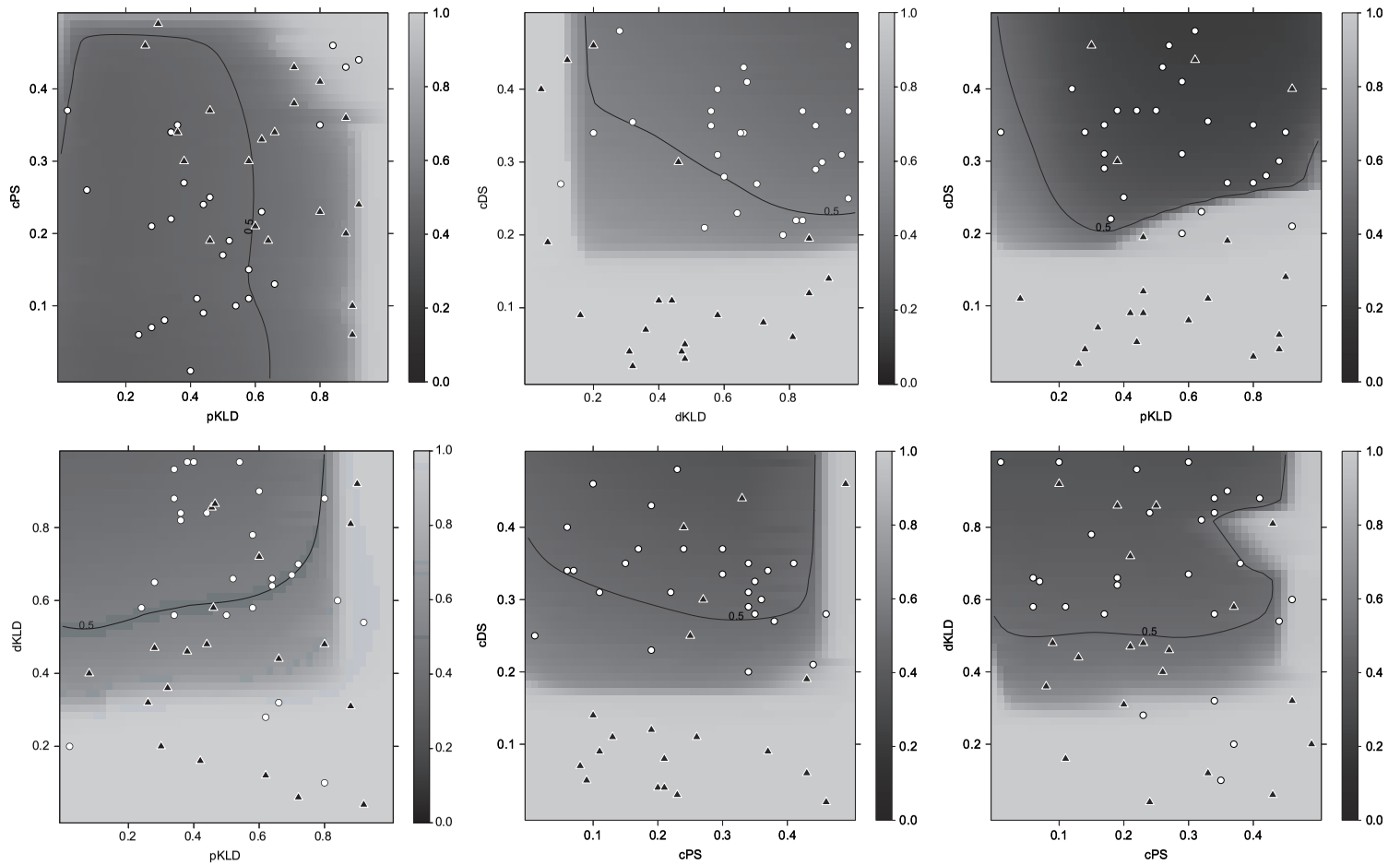

Fig. 2. Marginal predictive probability maps from the glioma data (each probability map is obtained by marginalizing over the two summary statistics that do not appear in the *x*- and *y*-labels of the images): ■, smaller probability of patients' death before 1 year (i.e. $\pi(Z_j = 1|\mathbf{Z}_{(-j)}, \mathcal{Y}_{(-j)}, \mathcal{Y}_j)$); ———, demarcation of the marginal decision boundary (i.e. $\pi(Z_j = 1|\mathbf{Z}_{(-j)}, \mathcal{Y}_{(-j)}, \mathcal{Y}_j) = 0.5$)—if $\pi(Z_j = 1|\mathbf{Z}, \mathcal{Y}_{(-j)}, \mathcal{Y}_j) > 0.5$, we predict death of that patient before 1 year; ○, true 1-year survival status, patient lived longer than 1 year; ▲, true 1-year survival status, patient died before 1 year

100000 iterations. In each iteration, we oversample (10:1) draws from the posterior of the stage II parameters. Stage I takes around 20 h for all 47 patients, whereas stage II takes 5 min. The algorithm is programmed in C and implemented on a 3.0-GHz Mac Xserve computer. The first 50 000 draws are discarded as burn-in. By visual inspection of the trace plots of the (fixed dimension) parameters, the burn-in is sufficient and the chain is sampling from the posterior (stationary) distribution.

We calculate the cross-validated correct classification rate CCR_{cv}—the proportion of correctly predicted survival statuses—the positive predictive value PPV_{cv}—the probability of death within 1 year conditional on prediction of death within 1 year—and the negative predictive value NPV_{cv}—the probability of survival longer than 1 year given a prediction of survival longer than 1 year. The results are $CCR_{cv} = 0.787$ (37/47), $PPV_{cv} = 0.813$ (13/16) and $NPV_{cv} = 0.774$ (24/31) (Table 1, first row). Investigators are interested in therapy intervention or modification if the model accurately predicts death within 1 year. Therefore, PPV_{cv} is of greater interest than NPV_{cv}.

In Fig. 2 we display the six bivariate marginal predictive probability maps. On each map is the marginal decision boundary separating the space of covariates into two regions based on whether $\pi(Z_{new} = 1|\mathbf{Z}, \mathbf{X}_{new}) > 0.5$. Also shown in Fig. 2 are the posterior means of the covariates for all 47 subjects. The triangles represent those subjects who died before 1 year, and the circles represent those who lived longer than 1 year. The probabilities in the maps are $\pi(Z_{new} = 1|\mathbf{Z}, \mathbf{X}_{new})$ marginalized over the six combinations of pairs of covariates. It is evident that the marginal decision boundaries are quite complex. From Fig. 2 we see that, marginally, small values of dKLD and cDS are associated with poor survival and that large values of pKLD are also associated with poor survival. There are also substantial interactions between cPS and dKLD, between dKLD and cDS and between pKLD and dKLD. In general, the overall gross pattern of increases in the dKLD- and cDS-statistics are predictive of longer survival—consonant with what our colleagues hypothesized. However, the overall gross pattern of decreases in the pKLD- and cPS-statistics are predictive of shorter survival—dissonant with that hypothesized. One plausible explanation provided by our colleagues is that a reduction in RCBF creates a hypoxic environment within the tumour and hypoxia is known to be protective against radia-

Table 1. Model comparisons†

Model	CCR_{cv}‡	PPV_{cv}§	NPV_{cv}§§
Bayesian joint model	0.787 (37/47)	0.813 (13/16)	0.774 (24/31)
Separate models (two-stage model)	0.830 (39/47)	0.853 (15/18)	0.827 (24/29)
Separate models (stage I + probit)	0.617 (29/47)	0.572 (11/20)	0.667 (18/27)
Bayesian joint model	0.723 (34/47)	0.733 (11/15)	0.719 (23/32)
Single model (stage II only)	0.638 (30/47)	0.600 (9/15)	0.656 (21/32)

†Upper part, comparisons based on all four statistics: first row, our proposed joint model; second row, separate two-stage model (not joint) (the second-stage model (GNLM–BMARS) runs conditional on the posterior expectations of the summary statistics from stage 1); third row, separate two-stage model (the second-stage model is a standard probit regression model); summary covariates are fixed at their posterior expectations from stage I. Bottom part, comparisons using only the two KLD-statistics: fourth row, our proposed model; fifth row, summary statistics computed on observed data (GNLM–BMARS runs conditional on 'observed' dKLD and pKLD).
‡Correct cross-validated classification rate.
§Cross-validated positive predictive value.
§§Cross-validated negative predictive value.

tion damage. However, we caution that the exact mechanism is unknown and that it warrants further investigation (Galbán *et al.*, 2009).

The baseline prognostic factors age, type of surgery, Karnofsky performance score, pathology grade and tumour size were also included in stage II as covariates. However, their inclusion did not increase the overall correct classification rate and each was included in the model less than 20% of the time (either as main effects or in an interaction term).

Each of the four summary statistics were included in the joint model as either a main effect or as an interaction term a high percentage of the MCMC draws (dKLD, 95.9%; pKLD, 90.5%; cDS, 81.1%; cPS, 84.3%). This indicates their importance in predicting survival. Both dKLD and pKLD appear to be slightly stronger predictors than either cDS or cPS on the basis of the amount of time spent in the model.

4. Model assessment

4.1. Comparison with simpler models

Our first comparison is with two separate models (i.e. not modelled jointly). The image data are fitted with our multivariate pairwise difference prior model. The posterior means of the summary statistics are treated as fixed known values and used as covariates in our GNLM–BMARS model. Thus the only difference between this procedure and our joint model is that our joint model accounts for the uncertainty in the summary statistics. Using point estimates, such as the posterior means, of the summary statistics as covariates in stage II results in overly optimistic prediction errors (see, for example, Little and Rubin (2002)). Ignoring the sampling variability in stage I, two additional patients are correctly classified (Table 1, second row). The posterior means of the summary statistics for these patients are near the decision boundary and happened to fall on the correct side, whereas the maximum *a posteriori* probability estimate was on the other side. Accounting for the variability in these random statistics, therefore, is necessary for robust prediction.

Our second comparison is again with two separate models. We estimate the posterior means of the summary statistics from our multivariate pairwise difference prior model. These point estimates are then treated as fixed known covariates and put into a standard probit regression model (Table 1, third row). Both main effects and interaction terms are allowed in the probit regression model. The Bayesian information criterion is used for model selection. Correct prediction from our joint model is much higher even though uncertainty in the covariates is ignored, as well as model uncertainty, in the separate multivariate pairwise difference prior plus probit regression model. The extra flexibility that is afforded by the BMARS basis has a large effect on prediction.

Our final comparison illustrates the benefits of the spatiotemporal modelling in stage I by comparing our results with those based on the observed images. Since the cDS- and cPS-statistics rely on the conditional distribution of tumour response under the null hypothesis, it is not possible to derive these summary statistics on the observed images as we have no model to use to predict tumour null response. Thus, this comparison uses only dKLD and pKLD. We estimate dKLD and pKLD using the observed images by calculating the Kullback–Leibler divergence between observed tumour response and observed healthy tissue response (in the contralateral hemisphere) at week 3 and plugging these statistics into our GNLM–BMARS model as fixed covariates. Cross-validation results are shown in the bottom half of Table 1, fourth and fifth rows. Spatiotemporal modelling results in higher CCR_{cv}, PPV_{cv} and NPV_{cv}.

Our overall conclusion from these comparisons is that joint modelling of the spatiotempo-

ral structure in the images and the complexity in the decision boundary that is afforded by the covariates entering the GNLM model non-linearly and interacting in a complex manner is warranted for this data set. The images have complex structure and there is a complex relationship between the image-based summary statistics and 1-year survival. Furthermore, accounting for the uncertainty in the summary statistics and model averaging are necessary for robust prediction.

Results from simulation studies and sensitivity analyses can be found in the SWBM.

5. Discussion

In this paper, we propose a Bayesian joint model to predict early treatment efficacy based on QMRI data from patients with high grade gliomas. In stage I, we model the spatiotemporal structure in the QMRI data via a multivariate pairwise difference prior model and derive summary statistics as functionals of the parameters in the posterior. In stage II, a GNLM is used to classify each patient's 1-year survival status with the summary statistics derived in stage I as random predictors. The final predictive power is evaluated by cross-validation. Compared with previous work, our proposed joint model integrates many of the ideas that have been previously discussed. First, we extend the idea of the pairwise difference model to a multivariate setting, and, in fact, use it in a full spatiotemporal setting. The multivariate pairwise difference prior model accounts for spatiotemporal correlation in the images as well as the correlation between the diffusion and perfusion images. This results in an increase in the signal-to-noise ratio. Furthermore, data dimension reduction is realized by defining subject level summary statistics. Second, by utilizing the BMARS basis, we allow a flexible and complex classification model that can achieve high predictive power. Finally, our model accounts for the uncertainty in the stage II covariates and for the uncertainty in model selection, resulting in more robust predictions.

In our analysis, we dichotomize each patient's survival status at 1 year. However, there may not be any substantial difference between a patient who dies 11 months after diagnosis and a patient who dies at 13 months. Moreover, the censoring rate in the data is about 30% with a median follow-up of 23.1 months and all censored observations are greater than 1 year. Censoring may also play a role in the evaluation of tumour treatment efficacy and dichotomizing survival may lead to inefficient estimation. We are currently building a joint imaging–survival model, where, in stage II, we model the censored survival times explicitly.

We note here that we propose four summary statistics that capture information about the early changes in ADC and RCBF due to treatment. Results show that they perform well in terms of good prediction. We do not claim that these summary statistics capture the most, or even the best, information. Information is always lost in data reduction. Much more research is needed to determine how much data reduction is tolerable. Reduction to four summary statistics does a good job, but perhaps five or six would be better. We did not attempt to use more than four summary statistics owing to the limited sample size in the pilot study. With larger sample sizes, less data reduction may be beneficial.

Our results show that early changes in diffusion and perfusion appear to be valuable biomarkers for the early assessment of treatment efficacy. These result are promising, albeit preliminary. The ability to predict treatment response during therapy, as opposed to waiting to assess traditional radiologic response, has the potential to facilitate patient management and may allow second-line or salvage therapies to begin earlier than current practice dictates. Lastly, our model and sampling algorithm are easily extendable to more than two image types at more than two times points with more than four summary statistics.

Acknowledgements

We thank the Joint Editor and the referees for their helpful comments and suggestions which led to substantial improvement of the original manuscript. This work was funded by the US National Institutes of Health grant P01 CA085878.

References

Albert, J. H. and Chib, S. (1993) Bayesian analysis of binary and polychotomous response data. *J. Am. Statist. Ass.*, **88**, 669–679.

Basser, P. J. and Jones, D. K. (2002) Diffusion-tensor MRI: theory, experimental design and data analysis—a technical review. *NMR Biomed.*, **15**, 456–467.

Besag, J. (1993) Towards Bayesian image analysis. *J. Appl. Statist.*, **20**, 107–119.

Chenevert, T. L., Stegman, L. D., Taylor, J. M. G., Robertson, P. L., Greenberg, H. S., Rehemtulla, A. and Ross, B. D. (2000) Diffusion magnetic resonance imaging: an early surrogate marker of therapeutic efficacy in brain tumors. *J. Natn. Cancer Inst.*, **92**, 2029–2036.

Chinnaiyan, A. M., Prasad, U., Shankar, S., Hamstra, D. A., Shanaiah, M., Chenevert, T. L., Ross, B. D. and Rehemtulla, A. (2000) Combined effect of tumor necrosis factor-related apoptosis-inducing ligand and ionizing radiation in breast cancer therapy. *Proc. Natn. Acad. Sci. USA*, **97**, 1754–1759.

Denison, D. G. T., Holmes, C. C., Mallick, B. K. and Smith, A. F. M. (2002) *Bayesian Methods for Nonlinear Classification and Regression*. Chichester: Wiley.

Friedman, J. H. (1991) Multivariate adaptive regression splines. *Ann. Statist.*, **19**, 1–61.

Galbán, C. J., Chenevert, T. L., Meyer, C. R., Tsien, C., Lawrence, T. S., Hamstra, D. A., Junck, L., Sundgren, P. C., Johnson, T. D., Ross, D. J. and Chenevert, T. L. (2009) The parametric response map is an imaging biomarker for early cancer treatment outcome. *Nat. Med.*, **15**, 572–576.

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

Green, P. J. (1995) Reversible jump Markov chain Monte Carlo computation and Bayesian model determination. *Biometrika*, **82**, 711–732.

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 .

Hamstra, D. A., Galbán, C. J., Meyer, C. R., Johnson, T. D., Sundgren, P. C., Tsien, C., Lawrence, T. S., Junck, L., Ross, D. J., Rehemtulla, A., Ross, B. D. and Chenevert, T. L. (2008) Functional diffusion map as an early imaging biomarker for high-grade glioma: correlation with conventional radiologic response and overall survival. *J. Clin. Oncol.*, **26**, 3387–3394.

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

Holmes, C. C. and Denison, D. G. T. (2003) Classification with Bayesian MARS. *Mach. Learn.*, **50**, 159–173.

Kullback, S. and Leibler, R. A. (1951) On information and sufficiency. *Ann. Math. Statist.*, **22**, 79–86.

Law, M., Young, R., Babb, J., Pollack, E. and Johnson, G. (2007) Histogram analysis versus region of interest analysis of dynamic susceptibility contrast perfusion MR imaging data in the grading of cerebral gliomas. *Am. J. Neurorad.*, **28**, 761–766.

Laws, E. R., Parney, I. F., Huang, W., Anderson, F., Morris, A. M., Asher, A., Lillehei, K. O., Bernstein, M., Brem, H., Sloan, A., Berger, M. S., Chang, S. and Gliomas Outcome Investigators (2003) Survival following surgery and prognostic factors for recently diagnosed malignant glioma: data from the Glioma Outcomes Project. *J. Neurosurg.*, **99**, 467–473.

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

Liang, Z. P. and Lauterbur, P. C. (2000) *Principles of Magnetic Resonance Imaging: a Signal Processing Perspective*. New York: Wiley–Institute of Electrical and Electronics Engineers Press.

Little, R. J. A. and Rubin, D. B. (2002) *Statistical Analysis with Missing Data*. New York: Wiley-Interscience.

Meyer, C. R., Boes, J. L., Kim, B., Bland, P. H., Zasadny, K. R., Kison, P. V., Koral, K., Frey, K. A. and Wahl, R. L. (1997) Demonstration of accuracy and clinical versatility of mutual information for automatic multimodality image fusion using affine and thin-plate spline warped geometric deformations. *Med. Im. Anal.*, **1**, 195–206.

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.

Moffat, B. A., Chenevert, T. L., Meyer, C. R., McKeever, P. E., Hall, D. E., Hoff, B. A., Johnson, T. D., Rehemtulla, A.

and Ross, B. D. (2006) The functional diffusion map: an imaging biomarker for the early prediction of cancer treatment outcome. *Neoplasia*, **8**, 259–267.

Raftery, A. E., Madigan, D. and Volinsky, C. T. (1996) Accounting for model uncertainty in survival analysis improves predictive performance. *Baysn Statist.*, **5**, 323–349.

Ross, B. D., Chenevert, T. L., Kim, B. and Ben-Yoseph, O. (1994) Magnetic resonance imaging and spectroscopy: application to experimental neuro-oncology. *Quant. Nucl. Magn. Reson. Med. Biol.*, **1**, 89–106.

Young, R., Babb, J., Law, M., Pollack, E. and Johnson, G. (2007) Comparison of region-of-interest analysis with three different histogram analysis methods in the determination of perfusion metrics in patients with brain gliomas. *J. Magn. Reson. Imgng*, **26**, 1053–1063.

Zhao, M., Pipe, J. G., Bonnett, J. and Evelhoch, J. L. (1996) Early detection of treatment response by diffusion-weighted 1 H-NMR spectroscopy in a murine tumour in vivo. *Br. J. Cancer*, **73**, 61–64.

Supporting information

Additional 'supporting-information' may be found in the on-line version of this article:

'Supporting information for Predicting treatment efficacy via quantitative MRI: a Bayesian joint model'.

Please note: Wiley–Blackwell are not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the author for correspondence for the article.