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

Year 2007 *Paper* 73

Bayesian Bivariate Image Analysis with Application to Dual Autoradiography

Timothy D. Johnson* Morand Piert†

*University of Michigan, tdjtdj@umich.edu
†University of Michigan

Bayesian Bivariate Image Analysis with Application to Dual Autoradiography

Timothy D. Johnson and Morand Piert

Abstract

We present a Bayesian bivariate image model and apply it to a study that was designed to investigate the relationship between hypoxia and angiogenesis in an animal tumor model. Two radiolabeled tracers (one measuring angio- genesis, the other measuring hypoxia) were simultaneously injected into the animals, the tumors removed and autoradiographic images of the tracer concentrations were obtained. We model correlation between tracers with a mixture of bivariate normal distributions and the spatial correlation inherent in the images by means of the celebrated Potts model. Although the Potts model is typically used for image segmentation, we use it solely as a device to account for spatial correlation. The number of classes in the model is assumed un- known and is estimated via reversible jump MCMC, marginalizing over the number of classes for posterior inference. We present the model and estimation method using set theory notation which will assist us in introducing a novel reallocation scheme used in the reversible jump proposals. We also estimate the spatial regularization parameter in the Potts model prior. Through a simulation study, we show that it is necessary to account for both the spatial correlation and the correlation between the two tracers.

Bayesian Bivariate Image Analysis with Application to Dual Autoradiography

Timothy D. Johnson *
University of Michigan
Department of Biostatistics
School of Public Health

Morand Piert
University of Michigan
Department of Radiology
School of Medicine

We present a Bayesian bivariate image model and apply it to a study that was designed to investigate the relationship between hypoxia and angiogenesis in an animal tumor model. Two radiolabeled tracers (one measuring angiogenesis, the other measuring hypoxia) were simultaneously injected into the animals, the tumors removed and autoradiographic images of the tracer concentrations were obtained. We model correlation between tracers with a mixture of bivariate normal distributions and the spatial correlation inherent in the images by means of the celebrated Potts model. Although the Potts model is typically used for image segmentation, we use it solely as a device to account for spatial correlation. The number of classes in the model is assumed unknown and is estimated via reversible jump MCMC, marginalizing over the number of classes for posterior inference. We present the model and estimation method using set theory notation which will assist us in introducing a novel reallocation scheme used in the reversible jump proposals. We also estimate the spatial regularization parameter in the Potts model prior. Through a simulation study, we show that it is necessary to account for both the spatial correlation and the correlation between the two tracers.

Key Words: Autoradiography; Bayesian analysis; Image analysis; Markov random fields; RJMCMC; Swendsen-Wang algorithm.

*corresponding author, email: tdjtdj@umich.edu;
address: 1420 Washington Heights, Ann Arbor, MI 48109–2029

1 Introduction

This work is motivated by a study investigating the relationship between hypoxia and angiogenesis in an animal tumor model. Oxygen deficiency (hypoxia) is a common feature of malignant tumors and has the well known effect of decreasing the sensitivity of tumors to ionizing radiation. It has been identified as a factor for tumor progression and for resistance to anticancer treatment. The proliferation of new vessels from pre-existing capillaries (angiogenesis) is a key player in the pathological development of solid tumors and for their ability to metastasize. It has long been conjectured that as solid tumors grow, the central core of the tumor becomes necrotic due to hypoxia and that this core of hypoxic tissue is surrounded by an annulus of tissue with high angiogenic activity. With the recent development of in vivo tracers for hypoxia and angiogenesis, researchers are now able objectively study the relationship between hypoxia and angiogenesis within tumors.

The study consisted of eleven Swiss nude mice. Each mouse received two xenografts of EMT6 tumors approximately $1\,\mathrm{mm}^3$ to both sides of the thoracic back. Two weeks later, the mice were injected with two radio-labeled tracers: [^{18}F]FAZA and [^{125}I]gluco-RGD. [^{18}F]FAZA ($t_{1/2}$ = 109.7 min.) measures hypoxia (Piert et al., 2005) and [^{125}I]gluco-RGD ($t_{1/2}$ = 59.4 days) measures angiogenesis (Haubner et al., 1999). Three hours later the animals were euthanized, their tumors dissected, frozen and cut into 20 μm sections. Immediately after sectioning, digital autoradiography was performed for one hour. After complete decay of ^{18}F, digital autoradiography was again performed for one hour. The resulting image—representing ^{125}I activity—was subtracted from the first image resulting in the "true" ^{18}F distribution. Digital autoradiography was again performed for 24 hours to capture the ^{18}I distribution. To account for the disparate exposure times, tracer activity in the tumor was normalized by dividing it by tracer activity in healthy muscle tissue; after subtracting background activity from both tumor and muscle. Henceforth the [^{18}F]FAZA to muscle ratio will be denote "FAZA" and the [^{125}I]gluco-RGD to muscle ratio will be denoted "RGD". Full

details of the study will appear in a companion manuscript. The purpose of this article is to present, in detail, the imaging model used to study the spatial relationship between hypoxia and angiogenesis.

We present a Bayesian model that accounts for both the spatial correlation (the spatial distribution of the tracers) as well as the correlation between tracers. To our knowledge, this is the first, fully Bayesian, imaging model has been applied to dual autoradiographic images. We model the correlation between tracers with a mixture of bivariate normal distributions. The spatial correlation is modeled by a hidden Markov random field. The number of classes in the mixture is assumed to be unknown and is estimated via reversible jump Markov chain Monte Carlo (RJMCMC) simulation (Green, 1995). Our main interest is not in directly estimating the number of classes, nor in segmentation, but rather in the distributional relationship between the two tracers. Thus, we marginalize over the number of classes in our posterior inference. We also employ the Swendsen-Wang algorithm (Swendsen and Wang, 1987). This algorithm has been shown to have superior mixing properties as opposed to the commonly used full conditional, pixel-wise, Gibbs updates of class membership. Results are sensitive to the regularization parameter used in our prior model, thus this parameter is estimated from the data (however, it is commonly assumed know).

Image analysis crosses many fields of study including statistics, engineering and computer science, to name a few. As a result, a variety of terminology has been used in the literature. To clarify this exposition and make it accessible to a wide audience, we present the model and estimation procedure using set theory ideas and notation, which should be familiar to most researchers across the various fields of study. This notation will also aid in describing our novel method for reallocating pixels to classes: which is necessary for the RJMCMC proposals.

We begin in Section 2 by introducing the model and notation. Section 3 is devoted to detailing our estimation of the posterior, including a description of the Swendsen-Wang

algorithm and the RJMCMC algorithm. Results from the motivating example are given in Section 4. A simulation study and sensitivity analysis are presented in the penultimate section. We wrap up the paper with a short discussion.

2 Bayesian bivariate image model

Our model is a Bayesian hierarchical model. Wherever possible, we use the language of set theory to describe elements of our model. All the set theory we use can be found in Halmos (1998). Let $\mathcal{P}$ denote the set of pixels in the image with $N = |\mathcal{P}|$. We subscript pixels by a single index, $i = 1, 2, \ldots, N$. Let $\mathbb{N}_K = \{1, 2, \ldots, K\}$. We suppose there is a finite hidden Markov random field (Besag, 1974) defined on $\mathcal{P}$ with finite state space $\mathcal{S}_K = \mathbb{N}_K$. Let $\mathbf{Z}^{\mathrm{T}} = (Z_1, Z_2, \ldots, Z_N)$ denote the image of latent states with $Z_i \in \mathcal{S}_K$ for all $i = 1, \ldots, N$. Define a relation on $\mathcal{P}$ as follows:

Definition 1 *Pixels i and j are related if $Z_i = Z_j$. Call this relation R. If i and j are related, we write iRj.*

Proposition 1 *R is an equivalence relation.*

Corollary 1 **Z** *partitions $\mathcal{P}$ via the equivalence relation R.*

The proofs of these follow from the definition. Henceforth, we refer to **Z** as the map of (equivalence) class labels with $Z_i \in \mathcal{S}_K \; \forall \, i \in \mathcal{P}$.

Definition 2 *Suppose* **Z** *partitions $\mathcal{P}$ into K equivalence classes. Pick one pixel from each equivalence class. Suppose these are $i_1, \ldots, i_K$. Then the kth equivalence class is denoted by $\mathcal{E}_{Rk}$, with $\mathcal{E}_{Rk} \equiv \{i \in \mathcal{P} : iRi_k\}$.*

Definition 3 *If pixel i and pixel j are adjacent to one another in the vertical or horizontal direction, then we say pixel i and pixel j are neighbors, denoted $i \sim j$.*

The prior for the class label map, $\mathbf{Z}$, is the celebrated Potts distribution (Potts, 1952):

$$\Pr(\mathbf{Z} = \mathbf{z} \mid \beta, K) = \exp\left\{-\beta \sum_{i \sim j} [1 - \delta_j(\mathcal{E}_{Ri})]\right\} \Big/ g(\beta,\, K) \tag{1}$$

with $\delta_x(A)$ denoting the Dirac measure that equals 1 if $x \in A$ and equals 0 if $x \notin A$ (i.e. $\delta_j(\mathcal{E}_{Ri}) = 1$ if and only if $Z_i = Z_j$). This distribution depends on a spatial regularization parameter, $\beta \geq 0$, and the number of equivalence classes, K. The norm alizing constant is

$$g(\beta,\, K) = \sum_{\mathbf{z} \in \mathcal{Z}_K} \exp\left\{-\beta \sum_{i \sim j} [1 - \delta_j(\mathcal{E}_{Ri})]\right\} \tag{2}$$

where $\mathcal{Z}_K = \{\mathbf{z} : z_i \in \mathcal{S}_K \; \forall\, i \in \mathcal{P}\}$: the set of all possible class label maps consisting of K classes. Note that $|\mathcal{Z}_K| = K^N$. Thus, even for moderate sized images and classes, the normalizing constant is computationally intractable. For example, if $K = 2$ and $N = 256$ (a small 16×16 image), then $|\mathcal{Z}_K| \approx 1.16 \times 10^{77}$. We treat both the normalizing constant and the number of classes as parameters, and so the normalizing constant must be estimated. In Section 3.2 we show how this can be done.

For $i \in \mathcal{S}_K$ let $\mathcal{N}_i = \{j \in \mathcal{S}_K : j \sim i\}$: the set of neighbors of pixel i. Suppose there are K equivalence classes. The Potts distribution (1) favors class label maps where neighbors are members of the same equivalence class. This is most easily seen by considering the conditional distribution of a single pixel, i;

$$\begin{aligned} \Pr(Z_i = k \mid \mathbf{Z}_{-i},\, \beta,\, K) &= \Pr(Z_i = k \mid \{Z_j : j \in \mathcal{N}_i\},\, \beta,\, K) \\ &= \exp\left\{-\beta \sum_{j \in \mathcal{N}_i} [1 - \delta_j(\mathcal{E}_{Rk})]\right\} \Big/ \sum_{\ell \in \mathcal{S}_K} \exp\left\{-\beta \sum_{j \in \mathcal{N}_i} [1 - \delta_j(\mathcal{E}_{R\ell})]\right\}. \end{aligned} \tag{3}$$

For a fixed β, this conditional probability is proportional to 1 if jRi, $\forall\, j \in \mathcal{N}_i$; is proportional to $\exp(-\beta)$ if jRi for exactly three $j \in \mathcal{N}_i$; is proportional to $\exp(-2\beta)$ if jRi for exactly two $j \in \mathcal{N}_i$; is proportional to $\exp(-3\beta)$ if jRi for exactly one $j \in \mathcal{N}_i$; and is proportional to $\exp(-4\beta)$ if jRi for zero $j \in \mathcal{N}_i$. If $\beta > 0$, we see that the Potts distribution favors class label maps where neighbors are members of the same equivalence class. As $\beta \to \infty$, the probability of neighbors belonging to the same equivalence class increases. When $\beta = 0$,

$\Pr(Z_i = k \mid \{Z_j : j \in \mathcal{N}_i\}, \beta, K) = K^{-1}$ for all $k \in \mathcal{S}_K$ which is independent of $\{Z_j : j \in \mathcal{N}_i\}$ (all pixels are mutually independent).

Let $\mathbf{Y}_F^{\mathrm{T}} = (Y_{1F}, \ldots, Y_{NF})$ denote the FAZA image and $\mathbf{Y}_R$ denote the RGD image. Define $\mathbf{Y} = (\mathbf{Y}_F^{\mathrm{T}}, \mathbf{Y}_R^{\mathrm{T}})$ and $\mathbf{Y}_i^{\mathrm{T}} = (Y_{iF}, Y_{iR})$. Given $\mathbf{Z} = \mathbf{z}$, we model $\mathbf{Y}_i$ as conditionally independent bivariate normal random variates: $\mathbf{Y}_i \mid z_i, \boldsymbol{\mu}_{z_i}, \Sigma_{z_i} \sim N(\boldsymbol{\mu}_{z_i}, \Sigma_{z_i})$. Thus,

$$f(\mathbf{y} \mid \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) = (2\pi)^{-N} \prod_{i=1}^{N} |\Sigma_{z_i}^{-1}| \exp[-0.5(\mathbf{y}_i - \boldsymbol{\mu}_{z_i})^{\mathrm{T}} \Sigma_{z_i}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_{z_i})],$$

where $\boldsymbol{\mu} = \{\boldsymbol{\mu}_k\}_{k \in \mathcal{S}_K}$ and $\boldsymbol{\Sigma} = \{\Sigma_k\}_{k \in \mathcal{S}_K}$. The priors for $\boldsymbol{\mu_k}$ and Σ_k^{-1}, $k \in \mathcal{S}_K$, are

$$\begin{aligned} \boldsymbol{\mu}_k &\overset{\text{iid}}{\sim} U([\min(\mathbf{Y}_F), \max(\mathbf{Y}_F)] \times [\min(\mathbf{Y}_R), \max(\mathbf{Y}_R)]) \\ \Sigma_k^{-1} &\overset{\text{iid}}{\sim} W(\mathrm{S}^{-1}, \nu_1) \\ \mathrm{S} &\sim W(\mathrm{T}, \nu_2) \end{aligned}$$

where $W(A, \nu)$ is the Wishart distribution with symmetric positive definite (SPD) scale matrix A and ν degrees of freedom and T is a diagonal matrix with $T_{11} = (\nu_1 - 3)(R_F/\nu_2)^2$ and $T_{22} = (\nu_1 - 3)(R_R/\nu_2)^2$ and R_F and R_R represent the range of FAZA and RGD, respectively. Here we chose $\nu_1 = \nu_2 = 10$.

A priori, $\beta \sim G(3, 2)$. The number of equivalence classes, M, is given a truncated Poisson distribution:

$$\Pr(M = K \mid \lambda) = \frac{\lambda^K}{K!} \left(\sum_{k=1}^{40} \lambda^j / j! \right)^{-1} I(K \in \{1, 2, \ldots, 40\}).$$

We set $\lambda = 15$.

Our model accounts for spatial correlation via the Potts prior on $\mathbf{Z}$ and for correlation between tracers via the bivariate normal distribution on $\mathbf{Y}$.

3 Posterior estimation

In the following subsections we detail the elements needed for posterior estimation. In particular we present details of the Swendsen-Wang algorithm (Swendsen and Wang, 1987), which

is used to efficiently update the map of equivalence class labels, $\mathbf{Z}$, and of the RJMCMC algorithm. The Swendsen-Wang algorithm can also be used to estimate ratios of the normalizing constants in equation (2). These ratios are necessary for the RJMCMC proposals and for updating the regularization parameter, β.

3.1 The Swendsen-Wang algorithm

The Swendsen-Wang algorithm (Swendsen and Wang, 1987) is a special case of slice sampling (Robert and Casella, 2005). The posterior distribution of $\mathbf{Z}$ given the data and all other parameters is

$$\Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{Y},\, \boldsymbol{\mu},\, \boldsymbol{\Sigma},\, \beta,\, K) \propto f(\mathbf{y} \mid \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \times \exp\left\{-\beta \sum_{i \sim j} \left[1 - \delta_j(\mathcal{E}_{Ri})\right]\right\}. \tag{4}$$

Now for each neighbor pair, $i \sim j$, of pixels define a real, non-negative random variable, V_{ij}. Let $\mathbf{V} = \{V_{ij}\}_{i \sim j}$ and define the conditional distribution of all V_{ij} given $\mathbf{Z}$ to be uniform and independent:

$$\pi(V_{ij} \mid \mathbf{Z}) = \exp\{\beta\left[1 - \delta_j(\mathcal{E}_{Ri})\right]\} I(0 \leq V_{ij} \leq \exp\{-\beta\left[1 - \delta_j(\mathcal{E}_{Ri})\right]\}). \tag{5}$$

Therefore,

$$\pi(\mathbf{V} \mid \mathbf{Z}) = \prod_{i \sim j} \exp\{\beta\left[1 - \delta_j(\mathcal{E}_{Ri})\right]\} I(0 \leq V_{ij} \leq \exp\{-\beta\left[1 - \delta_j(\mathcal{E}_{Ri})\right]\}). \tag{6}$$

Then it is easy to show that

$$\begin{aligned} &\Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{V},\, \mathbf{Y},\, \boldsymbol{\mu},\, \boldsymbol{\Sigma},\, \beta,\, K) \propto \\ &\qquad f(\mathbf{y} \mid \mathbf{Z}, \boldsymbol{\mu}, \boldsymbol{\Sigma}) \prod_{i \sim j} I(0 \leq v_{ij} \leq \exp\{-\beta\left[1 - \delta_j(\mathcal{E}_{Ri})\right]\}). \end{aligned} \tag{7}$$

We can therefore sample from the joint posterior distribution of $\mathbf{V}$ and $\mathbf{Z}$ by iterating between (6) and (7). Marginalizing over $\mathbf{V}$ we end up with the marginal posterior distribution of $\mathbf{Z}$ given in (4).

It is a simple matter to sample a $\mathbf{V} \mid \mathbf{Z}$ from (6). We simply draw each $V_{ij} \mid \mathbf{Z}$ from (5). To sample a new $\mathbf{Z}$ from (7) we note the following: for $\beta > 0$, using (5) we have

$$\Pr(V_{ij} > \exp(-\beta) \mid \mathbf{Z}) = \int_{\exp(-\beta)}^{\exp[-\beta(1-\delta_j(\mathcal{E}_{Ri})]} dV_{ij} > 0 \quad \text{iff} \quad \delta_j(\mathcal{E}_{Ri}) = 1 \quad \text{iff} \quad z_i = z_j. \tag{8}$$

That is, given that $V_{ij} > \exp(-\beta)$, z_i and z_j are constrained to be equal. We need the following definitions, a proposition and a corollary:

Definition 4 *Define a relation Q on $\mathcal{P}$ as follows: Pixels i and j are related (iQj) if there exists a set of pixels $\{i_\ell\}_{\ell=1}^{L} \subset \mathcal{P}$ such that $i = i_1$, $j = i_\ell$ and $V_{i_\ell, i_{\ell+1}} > \exp(-\beta)$ for $\ell = 1, \ldots, L-1$. If $L = 1$, then $V_{i_\ell, i_{\ell+1}} > \exp(-\beta)$ holds vacuously.*

Proposition 2 *Q is an equivalence relation.*

The proof of Proposition 2 is trivial and Q partitions $\mathcal{P}$.

Definition 5 *The partition defined by the equivalence relation Q on $\mathcal{P}$ is denoted by $\mathbf{Z}_Q$.*

Definition 6 *Suppose $\mathbf{Z}_Q$ partitions $\mathcal{P}$ into C equivalence classes. Pick one pixel from each equivalence class. Suppose these are $i_1, i_2, \ldots, i_C$. Then the cth equivalence class is denoted by $\mathcal{E}_{Qc}$. We note here that $\mathcal{E}_{Qc} \equiv \{i \in \mathcal{P} : iQi_c\}$.*

Corollary 2 *The partition $\mathbf{Z}_Q$ of $\mathcal{P}$ defined by Q is a refinement of the partition $\mathbf{Z}$ of $\mathcal{P}$ defined by R. That is if $i \in \mathcal{E}_{Qj}$ and $i \in \mathcal{E}_{Rk}$ then $\mathcal{E}_{Qj} \subseteq \mathcal{E}_{Rk}$.*

This corollary follows from the definitions of Q and R and from (8). As a direct consequence of Corollary 2 we have that $C \geq K$.

Now the left hand side of (7) can be factored as

$$\begin{aligned} \Pr(\mathbf{Z} = \mathbf{z} \mid \mathbf{V},\, \mathbf{Y},\, \boldsymbol{\mu},\, \boldsymbol{\Sigma},\, \beta,\, K) &= \prod_{c=1}^{C} \Pr(\{Z_i\}_{i \in \mathcal{E}_{Qc}} = z_{i_c} \mid \mathbf{V},\, \mathbf{Y},\, \boldsymbol{\mu},\, \boldsymbol{\Sigma},\, K) \\ &\times \prod_{i \sim j} I\{0 \leq v_{ij} \leq \exp[-\beta(1 - \delta_j(\mathcal{E}_{Ri})]\}, \end{aligned}$$

where $\{Z_i\}_{i \in \mathcal{E}_{Qc}} = z_{i_c}$ is shorthand notation for $Z_i = z_{i_c}$ for all $i \in \mathcal{E}_{Qc}$. Thus, each equivalence class can be updated independently of all other classes:

$$\Pr(\{Z_i\}_{i \in \mathcal{E}_{Qc}} = z_{i_c} \mid \mathbf{V},\, \mathbf{Y},\, \boldsymbol{\mu},\, \boldsymbol{\Sigma},\, K) \propto \prod_{i \in \mathcal{E}_{Qc}} |\Sigma_{z_{i_c}}^{-1}| \exp\left[-0.5(\mathbf{y}_i - \boldsymbol{\mu}_{z_{i_c}})^{\mathrm{T}} \Sigma_{z_{i_c}}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_{z_{i_c}})\right] \tag{9}$$

for $z_{i_c} \in S_K$. For the special case $\beta = 0$, condition (8) never holds: each pixel is its own equivalence class (relative to Q) and thus each pixel is updated independently of all other pixels.

3.2 Estimating ratios of the normalizing constant, $g(\beta,\, K)$

Evaluation of ratios of the normalizing constant, $g(\beta,\, K)$, is necessary to compute the prior ratio for the split and merge RJMCMC proposals and is required to draw values from the posterior distribution of the spatial regularization parameter, β. Recall that the normalizing constant is computationally intractable as it contains an astronomical number of summands. Hence we stochastically estimate it for $K = 1, \ldots, 40$ and on a grid of values for $\beta \in [0, 3]$ in increments of 0.01.

Stochastic estimation of the normalizing constant can be achieved efficiently by thermodynamic integration (Ogata, 1989) and use of the Swendsen-Wang algorithm. Estimation is performed prior to posterior simulation and results are stored in a matrix that can used during posterior simulation. Details are outlined in the Appendix.

3.3 RJMCMC

We assume that the number of hidden states (equivalently the number of equivalence classes), K, is unknown and estimate it via RJMCMC (Green, 1995; Richardson and Green, 1997; Green and Richardson, 2002). Associated with each equivalence class, $\mathcal{E}_{Rk}$, $k = 1, \ldots, K$, is a bivariate normal distribution with mean vector $\boldsymbol{\mu}_k$ and covariance matrix Σ_k. We define a pair of reversible jump moves: split and merge. Let $R \in \{s, m\}$ where s denotes the

splitting of a normal distribution into two normal distributions and m denotes a merging of two normal distributions into one. A split move requires the parameter space to increase in dimension by 5 (The mean of the bivariate normal distribution will be split into two new bivariate vector means, increasing the dimension by 2, and the covariance matrix will be split into two new covariate matrices, increasing the dimension by 3). A merge move decreases the dimension by 5. The probabilities of proposing these moves are

$$\begin{aligned} \Pr(R = s \mid K) &= c_k \min[1, \Pr(M = K+1 \mid \lambda)/\Pr(M = K \mid \lambda)] \\ \Pr(R = m \mid K+1) &= c_{k+1} \min[1, \Pr(M = K \mid \lambda)/\Pr(M = K+1 \mid \lambda)]. \end{aligned}$$

These probabilities are similar to those proposed by Green (1995). The c_ks are chosen so that $\Pr(R = s \mid K) + \Pr(R = m \mid K) = 1$.

3.3.1 Split proposal:

Suppose there are currently K classes. In the split move we propose to increase the state space from $\mathcal{S}_K$ to $\mathcal{S}_{K+1}$. A split move is proposed by randomly selecting a class to split: draw a $k \in \mathcal{S}_K$ uniformly (i.e. each with probability $1/K$). Suppose the class to be split is $\mathcal{E}_{Rk}$. Draw u_1 and u_2 independently from Beta(2, 2). Draw w_{11}, w_{12} and w_{22} independently from Beta(3, 3). Set $\mathbf{u}^{\mathrm{T}} = (u_1, u_2)$, $w_{21} = w_{12}$ and define W to be the 2×2 matrix with elements w_{ij}. Let $\mathbf{1}^{\mathrm{T}} = (1, 1)$. Dimension matching is performed via moment matching. Let $\Sigma_k^{1/2}$ denote the square-root of the SPD matrix Σ_k defined as follows: suppose the eigenvalues of Σ_k are λ_1, λ_2. Let $D = \mathrm{diag}\left(\sqrt{\lambda_1}, \sqrt{\lambda_2}\right)$. Let E denote a 2×2 matrix whose columns are eigenvectors of Σ_k. Then $\Sigma_k^{1/2} = EDE^{-1}$. Now the class is split into two new classes k^* and k^{**} by the following bijective transformation:

$$\begin{aligned} \boldsymbol{\mu}_{k^*} &= \boldsymbol{\mu}_k - \Sigma_k^{1/2}\mathbf{u} \\ \boldsymbol{\mu}_{k^{**}} &= \boldsymbol{\mu}_k + \Sigma_k^{1/2}\mathbf{u} \\ \Sigma_{k^*} &= 2\,W \odot \left[\Sigma_k^{1/2}(I - \mathbf{u}\mathbf{u}^{\mathrm{T}})\Sigma_k^{1/2}\right] \\ \Sigma_{k^{**}} &= 2\,(\mathbf{1}\mathbf{1}^{\mathrm{T}} - W) \odot \left[\Sigma_k^{1/2}(I - \mathbf{u}\mathbf{u}^{\mathrm{T}})\Sigma_k^{1/2}\right]. \end{aligned} \tag{10}$$

Here, $A \odot B$ denotes element-wise multiplication of the two conforming matrices, A and B. Now Σ_{k^*} and $\Sigma_{k^{**}}$ are not necessarily SPD. If either is not SPD, we reject the split proposal. Let $\boldsymbol{\mu}^{*\mathrm{T}}$ denote the new set of $K+1$ means and $\boldsymbol{\Sigma}^*$ denote the new set of $K+1$ covariance matrices.

Once an equivalence class is split into two new classes, pixels need to be reallocate to the new classes. To do so, we will use the machinery from the previous section. We stochastically reallocate all pixels from the current set of K equivalence classes to the new set of $K+1$ equivalence classes. We begin with the current partition $\mathbf{Z}$. Given $\mathbf{Z}$ we draw $\mathbf{V}^*$ using (6) with $\beta = \infty$. We define $\infty \cdot 0 = 0$ (anything times 0 equals 0). This defines a new equivalence relation Q^* and associated partition $\mathbf{Z}_Q^*$ with equivalence classes $\mathcal{E}_{Qc}^*$, $c = 1, \ldots, C^*$. We note here that for finite β, the partition $\mathbf{Z}_Q$ is stochastically determined via (5) and Definition 4. When $\beta = \infty$, the partition $\mathbf{Z}_Q^*$ is deterministically determined since (5) is an improper distribution with infinite mass at 0 when $\delta_j(\mathcal{E}_{Ri}) = 0$ (equivalently when $z_i \neq z_j$) and is uniformly distributed on $(0, 1)$ when $z_i = z_j$. Thus if $z_i = z_j$, $V_{ij}^* > \exp(-\infty)$ with probability 1. Now we draw $\mathbf{Z}^*$ with elements $z_i^* \in S_{K+1}$ using (9) with the appropriate changes:

$$\Pr(\{Z_i^*\}_{i \in \mathcal{E}_{Qc}^*} = z_{i_c}^* \mid \mathbf{V}^*, \mathbf{Y}, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, K) \propto \prod_{i \in \mathcal{E}_{Qc}^*} \left|\Sigma_{z_{i_c}^*}^{-1}\right| \exp\left[-0.5(\mathbf{y}_i - \boldsymbol{\mu}_{z_{i_c}^*})^{\mathrm{T}} \Sigma_{z_{i_c}^*}^{-1} (\mathbf{y}_i - \boldsymbol{\mu}_{z_{i_c}^*})\right]. \tag{11}$$

The probability of reallocation is given by

$$p_{\mathrm{alloc}}(\mathbf{z}^* \mid \mathbf{V}^*, \mathbf{Y}, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, K) = \prod_{c=1}^{C^*} \Pr(\{Z_i^*\}_{i \in \mathcal{E}_{Qc}^*} = z_{i_c}^* \mid \mathbf{V}^*, \mathbf{Y}, \boldsymbol{\mu}^*, \boldsymbol{\Sigma}^*, K), \tag{12}$$

while the probability of the current allocation of pixels (with elements $z_i \in S_K$) is given by

$$p_{\mathrm{alloc}}(\mathbf{z} \mid \mathbf{V}^*, \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, K) = \prod_{c=1}^{C^*} \Pr(\{Z_i\}_{i \in \mathcal{E}_{Qc}^*} = z_{i_c} \mid \mathbf{V}^*, \mathbf{Y}, \boldsymbol{\mu}, \boldsymbol{\Sigma}, K),$$

Note that $C^* \geq K$. Nevertheless, it is possible that when reallocating pixels to equivalence classes, we end up with less than $K+1$ classes. That is, one or more classes may be

empty due to the stochastic nature of the reallocation process. If this happens and the split move is accepted, the empty class(es) either get filled during a future update of the now current $\mathbf{Z}$ (with elements in $\mathcal{S}_{K+1}$) or is(are) merged with another class during a RJMCMC merge move. Our experience is that empty classes either get filled quickly or are remove quickly.

3.3.2 Merge proposal:

Suppose there are currently K classes. A merge move is attempted by randomly selecting an equivalence class $\mathcal{E}_{Ri}$, $i \in \mathcal{S}_K$, each with probability $1/K$. Suppose it is $\mathcal{E}_{Rk^*}$. Let $\boldsymbol{\mu}_{k^*}$ denote the mean and Σ_{k^*} denote the covariance matrix of the data corresponding to this equivalence class. The class to be merged with $\mathcal{E}_{Rk^*}$ is drawn at random from the remaining classes with probabilities

$$p_{\text{select}}(j \mid k^*) \propto 0.5 \left[\sqrt{(\boldsymbol{\mu}_{k^*} - \boldsymbol{\mu}_j)^{\mathrm{T}} \Sigma_j^{-1} (\boldsymbol{\mu}_{k^*} - \boldsymbol{\mu}_j)} + \sqrt{(\boldsymbol{\mu}_j - \boldsymbol{\mu}_{k^*})^{\mathrm{T}} \Sigma_{k^*}^{-1} (\boldsymbol{\mu}_j - \boldsymbol{\mu}_{k^*})} \right] \qquad \forall \quad j \in \mathcal{S}_K \setminus \{k^*\}$$

The right hand side is the average Mahalanobis distance of $\boldsymbol{\mu}_{k^*}$ to $N(\boldsymbol{\mu}_j, \Sigma_j)$ and of $\boldsymbol{\mu}_j$ to $N(\boldsymbol{\mu}_{k^*}, \Sigma_{k^*})$. Suppose the selected class is $\mathcal{E}_{Rk^{**}}$. The resulting class will be indexed by k with an appropriate relabeling of the other $K-1$ classes, as necessary. The inverse of the bijective transformation (10) is

$$\begin{aligned} \boldsymbol{\mu}_k &= 0.5\,(\boldsymbol{\mu}_{k^*} + \boldsymbol{\mu}_{k^{**}}) \\ \Sigma_k &= 0.5\,(\boldsymbol{\mu}_{k^*}\boldsymbol{\mu}_{k^*}^{\mathrm{T}} + \boldsymbol{\mu}_{k^{**}}\boldsymbol{\mu}_{k^{**}}^{\mathrm{T}} + \Sigma_{k^*} + \Sigma_{k^{**}}) - \boldsymbol{\mu}_k\boldsymbol{\mu}_k^{\mathrm{T}} \\ \mathbf{u} &= 0.5\,\Sigma_k^{-1/2}(\boldsymbol{\mu}_{k^{**}} - \boldsymbol{\mu}_{k^*}) \\ W &= 0.25\ (\Sigma_{k^*} - \Sigma_{k^{**}}) \oslash \left[\Sigma_k^{1/2}(I - \mathbf{u}\mathbf{u}^{\mathrm{T}})\Sigma_k^{1/2}\right] + 0.5\,\mathbf{1}\mathbf{1}^{\mathrm{T}} \end{aligned} \tag{13}$$

where $A \oslash B$ denotes element-wise division of the two conforming matrices A and B. The transformations (10) and (13) generalize those of Richardson and Green (1997) from the univariate to the bivariate normal case.

Next, we reallocate pixels to equivalence classes. We do so in the same manner as for the split proposal. We stochastically reallocate all pixels from the current set of K classes to the new set of $K-1$ classes. Given the current partition $\mathbf{Z}$, draw $\mathbf{V}^*$ using (6) with $\beta = \infty$. Now draw $\mathbf{Z}^*$ with elements $z_i^* \in S_{K-1}$ using (11). The probability of reallocation is given by (12) with these same changes.

Details of the acceptance probabilities for the split and merge move are given in the Appendix.

3.4 Updating β, $\boldsymbol{\mu}_k$, Σ_k^{-1} and S

β is updated via a Metropolis-Hasting step (Hastings, 1970). Gibbs steps are used to update S, $\boldsymbol{\mu}_k$, and Σ_k^{-1}, $k = 1, \ldots, K$. Let N_k denote the number of pixels in equivalence class k. The updates are

1. Draw $\beta^* \sim N(\beta, \sigma^2)$ subject to $\beta^* \in [0, 3]$ (If $\beta^* \notin [0, 3]$, reject the proposal). The proposal is accepted with probability $\min(1, A)$ where
$$A = \exp\left\{(\beta - \beta^*)\sum_{i\sim j}[1 - \delta_j(\mathcal{E}_{Ri})]\right\} g(\beta, K)/g(\beta^*, K)\exp[2(\beta - \beta^*)](\beta^*/\beta)^2.$$
2. Draw, for $k = 1, \ldots, K$,
$$\Sigma_k^{-1} \sim W\left[\left(S + \sum_{i\in\mathcal{E}_{Rk}}(\mathbf{y}_i - \boldsymbol{\mu}_k)(\mathbf{y}_i - \boldsymbol{\mu}_k)^{\mathrm{T}}\right)^{-1}, \nu_1 + N_k\right].$$
3. For $k = 1, \ldots, K$, draw $\boldsymbol{\mu}_k \sim N\left(\sum_{i\in\mathcal{E}_{Rk}}\mathbf{y}_i/N_k, \Sigma_k/N_k\right)$. Reject the draw if $\boldsymbol{\mu}_k \notin [\min(\mathbf{Y}_F), \max(\mathbf{Y}_F)] \times [\min(\mathbf{Y}_R), \max(\mathbf{Y}_R)]$.
4. Draw
$$S \sim W\left[\left(T^{-1} + \sum_{k=1}^{K}\Sigma_k^{-1}\right)^{-1}, \nu_2 + K\nu_1\right].$$

Following Higdon (1998), we alternate the Swendsen-Wang draw of $\mathbf{Z}$ with full conditional Gibbs updates of $Z_i \mid \mathbf{Z}_{-1}, \beta, K$ (via equation (3)) to help ensure movement within large equivalence classes, $\mathcal{E}_{Qj}$.

4 Results

We analyzed over 100 tumor slices from the eleven mice. (We were not able to register the tumors slices, build a 3D image of the tracer intensity and treat each tumor as a single image. There are no anatomical landmarks on the radiographs on which to register the slices and, for many of the tumors, the slices are not contiguous. Thus we analyzed each slice independently. However, it would not be difficult to extend our model to three dimensions.)

Of primary interest is the spatial distribution of hypoxia and angiogenesis. Unexpectedly we found that the conjecture noted in the introduction does not hold in general: these tumors are more heterogeneous than anticipated.

Two examples of the FAZA/RGD intensities are shown in the first row of Figures 1 and 3. The tumors are labeled "M1S6T1" and "M2S5T2", respectively in the figures. The marginal posterior mean images, $\widehat{E}(\boldsymbol{\mu} \mid Y)$, are displayed in the second row of these figures. The tumor displayed in Figure 1 shows, for the most part, the conjectured pattern of hypoxia in the core of the tumor surrounded by angiogenesis. In order to obtain a clearer picture of the pattern of hypoxia and angiogenesis, we partition the mean images into four distinct subsets of pixels. These are the set of all pixels where both FAZA and RGD tracer activity is high (intense hypoxia and angiogenesis); the set of all pixels where FAZA tracer activity is low and RGD activity is high; the set of all pixels where FAZA activity is high and RGD activity is low; and the set of all pixels where both tracer's activity is low. Thresholds need to be defined in order to precisely state what is meant by high or low tracer activity. For this purpose, we define tracer active as high (low) if the estimated posterior mean intensity is greater (less) than the empirical average intensity (averaged over all slices of all tumors). A pixel is classified into one of these four subsets if the posterior probability that both conditions hold is greater than 0.5. Precisely, these subsets are defined as

FAZA high/RGD high: $\{i \in \mathcal{P} : \Pr(\mu_{iF} > 3.18 \cap \mu_{iR} > 4.60 \mid \mathbf{Y}) > 0.5\}$. This set corresponds to higher than average hypoxia and angiogenesis.

FAZA low/RGD low: $\{i \in \mathcal{P} : \Pr(\mu_{iF} < 3.18 \cap \mu_{iR} < 4.60 \mid \mathbf{Y}) > 0.5\}$. This set corresponds to lower than average hypoxia and angiogenesis.

FAZA low/RGD high: $\{i \in \mathcal{P} : \Pr(\mu_{iF} < 3.18 \cap \mu_{iR} > 4.60 \mid \mathbf{Y}) > 0.5\}$ This set corresponds to lower than average hypoxia and higher than average angiogenesis.

FAZA high:RGD low: $\{i \in \mathcal{P} : \Pr(\mu_{iF} > 3.18 \cap \mu_{iR} < 4.60 \mid \mathbf{Y}) > 0.5\}$. This set corresponds to higher than average hypoxia and lower than average angiogenesis.

Pixels within the tumor boundaries that are not members of any of the above sets are colored in black. The values 3.18 and 4.60 are the empirical average intensities of FAZA and RGD. The third row of Figure 1 displays the partitioned image. From this image it is evident that hypoxia is larger than the average and angiogenesis is lower than average over most of the core of the tumor (colored yellow). Near the lower left and upper right regions of the tumor, angiogenic activity is high and hypoxia is low (colored red) . Neither hypoxia nor angiogenesis is present in the outer annulus of the tumor (colored blue). This is observed in nearly every tumor. We suspect this is an artifact caused by two factors. One, the partial volume effect (pixels near the boundary contain both healthy and diseased tissue). Two, imprecise outlining of the tumor boundary. Also of interest is the large area where both high levels of hypoxia and angiogenesis coexist (colored gray).

The example shown in Figure 3 tells a different story. Here, there is a region of highly angiogenic tissue appearing preferentially in the core of the tumor surrounded by a region of tissue that is largely hypoxic.

The harmonic mean of the posterior standard deviation of the tracers are displayed in Figures 2 and 4. Overall, both figures show that the variability is larger for RGD than it is for FAZA. Both images also show some evidence of negative correlation between tracers in the core of the tumors. The tracers become more positively correlated near the periphery.

Room does not permit us to show results from all 100+ slices. However, there was much heterogeneity across the tumors as well as across slices of a single tumor. Many tumors

show evidence of a core that is both poorly oxygenated and has high angiogenic activity. Some of these cores are surrounded by hypoxic tissue and others surrounded by tissue high in angiogenic activity. Others appear to be well oxygenated and lack angiogenic activity. No general summary of the spatial distribution of FAZA and RGD could be ascertained. These results call for further investigation. They also suggest that in vivo imaging of tracer uptake in tumors may be useful in tailoring cancer treatment to the individual. For example, tumors that show mostly angiogenic activity should be treated with an anti-angiogenic drug.

Overall, the RJMCMC algorithm performed quite well. Most of the autoradiographs that we analyzed had acceptance rates on the order of 5% to 18% for the split and merge moves, combined. However, there were a handful of cases that had acceptance rates as low as 0.5%. We believe part of the success of the RJMCMC acceptance rates is a result of the way we reallocate pixels to the new set of classes. Each equivalence class, $\mathcal{E}_{Rk}$, is refined to a new set of equivalence classes (via the relation Q^*), and the pixels in these new classes are reallocated independently of other classes. The components of the refinement tend to be much smaller than the original components. Thus the algorithm easily accepts new small components and easily merges these components. In fact, our first attempt was to reallocate pixels as proposed by Green and Richardson (2002). They update only those pixels in the class chosen to be split or the two classes chosen to be merged. Reallocation of pixels for a merge move is deterministic. For a split move, they perform a raster scan on those pixels to be reallocated to the two new classes. Each pixel being updated conditionally on the class value of its neighbors that have already been updated. For our application, performance of the reversible jump algorithm was extremely poor, using their reallocation strategy, and so we abandoned it in search of a better reallocation scheme for our application.

5 Simulation study and sensitivity to priors

5.1 Simulation study

We conduct a simulation study to assess the performance of our model against various models with simpler structure. In particular, we compare the bias of parameter estimates from our model, to our model with the number of classes fixed, a priori, at four; to a spatial model where the correlation between the two images is assumed to be zero; and to a bivariate model where the spatial structure of the data is ignored. We also compare the misclassification rates of these models. The classification is based on partitioning the images into the four subsets defined in the previous section. Further, we compare the empirical misclassification rate by assuming no model and determining which data points fall into the four subsets (here the subsets are based solely on the observed intensities at a pixel being larger (smaller) than their respective empirical averages).

We create twenty bivariate images by partitioning a 100×100 grid into sixteen 25×25 regions. For each image, two regions are selected at random, without replacement, and populated with draws from one of eight bivariate normal distributions. Parameter values for these distributions are listed at the bottom of Table 1. An example of one of these twenty data sets is shown in Figure 5.

Bias results from this simulation study are given in Table 1. From this table, we see that our proposed model is the least biased of the four. The non-spatial, bivariate model estimates the variances quite well, but does not estimate the means or the correlation very well. The other two models have quite large biases for all parameters.

Overall misclassification rate results are similar to the bias results. The misclassification rate is smallest for our model. In the simulation study, we assume that the empirical average threshold of the two tracers are both 0.4. The average misclassification rates are listed in Table 2. The classification results from the simulated data set shown in Figure 5 are displayed in Figure 6.

We also perform a similar study generating uncorrelated images. We compare our bivariate image model with the spatial model assuming an uncorrelated covariance structure. Overall, nothing substantive is lost use the bivariate image model. Bias results and misclassification rates (not shown) are quite similar between the two models.

5.2 Sensitivity analysis

For the RJMCMC algorithm, the number of parameters has been shown to be sensitive to the prior on the number of, in the present case, equivalence classes (Green,1995). To investigate this sensitivity, we change the prior on the number of equivalence classes. Recall that the number of equivalence classes has a truncated Poisson prior with parameter $\lambda = 15$. We investigate the sensitivity to this prior when we change λ to 10 and to 20. The resulting posterior distributions of the number of equivalence classes are displayed in Figure 7. Although the posterior of the number of equivalence classes is sensitive to its prior, the partitioned images are rather insensitive due in large part to model averaging. For the first data set, only 1.29% and 1.16% of the pixels are classified differently when λ changes from 15, to 10 and 20, respectively. For the second, only 1.47% and 0.79% of the pixels are classified differently when λ changes from 15, to 10 and 20, respectively.

6 Conclusion

We have developed a Bayesian bivariate image model that is useful in analyzing the spatial distribution of hypoxia and angiogenesis in an animal tumor model. The results from this study suggest tumors are more heterogeneous in their spatial distribution of angiogenic activity and hypoxia than previously thought. This also suggests that in vivo imaging of tumors can aid in the individualized treatment of solid tumors. Tumors that are largely angiogenic should be treated with anti-angiogenic agents, tumors that are largely hypoxic should be treated with hypoxia targeted treatments, while heterogeneous tumors should be

treated with a combination of treatments.

Posterior estimation takes about 60 minutes (50,000 samples from the posterior are used for estimation after a burn-in of 10,000) on a Macintosh Quad 2.5 GHz PowerPC G5 computer. The code was written in the C programming language. The dimension of each marginal image is 100×100. The down-side of our model and estimation procedure is the time it takes to estimate the log ratios of normalizing constants, $\lambda_K(a, b)$. There are 12,040 points on the 2D grid at which we estimate these ratios. It takes approximate 48 hours of CPU time to estimate these ratios. We note here that this the estimation was done once for a 100×100 bivariate image. Our estimation procedure is performed over the entire 100×100 image, including both tumor and non-tumor tissue. We do not compute the ratios for every individual tumor. After estimation, we mask out the non-tumor regions.

As a result of good mixing properties in the RJMCMC algorithm and model averaging, we are able to account for the variability due to the uncertainty in the model (that is, in the number of equivalence classes). As a result, this leads to improved parameter estimation and quantification of model uncertainty (Robert, 2001). The simulations studies suggest that the level of complexity in our model and estimation procedures are both necessary to reduce bias and to appropriately account for the correlation between the two tracers. Furthermore, little is lost if simpler models hold.

Appendix

Estimating the normalizing constant

The outline of the estimating procedure follows the general method found in Gelman and Meng (1998) with appropriate notational changes. For a given K, we wish to estimate the

log ratio of normalizing constants $\lambda_K(a,b) \equiv \ln[g(b,K)/g(a,K)]$ for $b > a \geq 0$. Now

$$\begin{aligned}
\frac{d\,\ln[g(\beta,K)]}{d\beta} &= \frac{d\,\ln\left(\sum_{\mathbf{z}\in\mathcal{Z}_K}\exp\left\{-\beta\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\}\right)}{d\beta} \\
&= \sum_{\mathbf{z}\in\mathcal{Z}_K}\frac{\frac{d}{d\beta}\exp\left\{-\beta\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\}}{\sum_{\mathbf{z}\in\mathcal{Z}_K}\exp\left\{-\beta\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\}} \times \frac{\exp\left\{-\beta\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\}}{\exp\left\{-\beta\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\}} \\
&= -\sum_{\mathbf{z}\in\mathcal{Z}_K}\left\{\Pr(\mathbf{Z}=\mathbf{z}\mid\beta,\,K)\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\} \\
&= -\mathrm{E}_\beta\left\{\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\}.
\end{aligned}$$

Therefore,

$$\lambda_K(a,b) = \int_a^b \frac{d\,\ln[g(\beta,K)]}{d\beta}d\beta = -\int_a^b \mathrm{E}_\beta\left\{\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\}d\beta. \tag{14}$$

In order to estimate $\lambda_K(a,b)$, we have two integrals to evaluate. The inner integral, or expectation, is estimated via MCMC. We use the Swendsen-Wang algorithm under the assumption that there is no likelihood term. That is, in (7) we set $f(\mathbf{y}\mid\mathbf{Z},\,\boldsymbol{\mu},\,\boldsymbol{\Sigma}) = 1$. Thus, (9) becomes $\Pr(\{Z_i\}_{i\in\mathcal{E}_{Qc}} = z_{i_c}\mid\mathbf{V},\,K) = K^{-1}$ for $z_{i_c}\in S_K$: each equivalence class is updated independently with equal probability that it takes one of the K class labels. After a burn-in of 1000 iterations we average $\sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]$ over the next 1000 iterations. The outer integral in equation (14) is evaluated numerically using the trapezoidal rule. The ratio is calculated for $K = 1,\ldots,40$ and $\beta = 0, 0.01,\ldots,3$. For values of a and b not on the grid on which the (inner) expectation is evaluated, we linearly interpolate.

For the RJMCMC proposal, the ratio $\ln[g(\beta,K)/g(\beta,K+1)]$ is required. This can be evaluated using the fact that $\ln[g(0,K)] = \ln\sum_{\mathbf{z}\in\mathcal{Z}_K}1 = N\ln K$ and the following identity:

$$\ln[g(\beta,\,K)/g(\beta,K+1)] = \lambda_K(0,\beta) - \lambda_{K+1}(0,\beta) + \ln[g(0,K)] + \ln[g(0,K+1)].$$

RJMCMC split and merge proposals

Let

$$\Sigma_k^{1/2} = \begin{bmatrix} \psi_{11} & \psi_{12} \\ \psi_{12} & \psi_{22} \end{bmatrix}.$$

Then the determinant of the Jacobian of the transformation (10) is

$$\begin{aligned} \det(J) &= 1024(\psi_{11}+\psi_{22})(\psi_{11}\psi_{22}-\psi_{12}^2)^2 \times \\ &\quad \{[\psi_{11}(u_1-1)+\psi_{12}u_2][\psi_{11}+\psi_{11}u_1+\psi_{12}u_2]-\psi_{12}^2\} \times \\ &\quad \{[\psi_{12}(u_1-1)+\psi_{22}u_2][\psi_{12}+\psi_{12}u_1+\psi_{22}u_2]-\psi_{22}^2\} \times \\ &\quad \{\psi_{11}\psi_{12}(u_1^2-1)+\psi_{11}\psi_{22}u_1u_2+\psi_{12}[\psi_{12}u_1u_2+\psi_{22}(u_2^2-1)]\}. \end{aligned}$$

Suppose there are currently K classes. The acceptance probability for a split proposal is given by $\min(1, A_s)$ where

$$A_s = R_{\text{prop}} \times R_{\text{prior}} \times R_{\text{likeli}} \times |\det(J)|\,.$$

R_{prop} is the proposal ratio, given by

$$\begin{aligned} R_{\text{prop}} &= \frac{c_{k+1}\,(K+1)}{c_k\,\lambda} \times \frac{K\,p_{\text{select}}(k^{**} \mid k^*)}{K+1} \times \frac{p_{\text{alloc}}(\mathbf{z} \mid \mathbf{V}^*,\, \mathbf{Y},\, \boldsymbol{\mu}^*,\, \boldsymbol{\Sigma}^*,\, K)}{p_{\text{alloc}}(\mathbf{z}^* \mid \mathbf{V}^*,\, \mathbf{Y},\, \boldsymbol{\mu}^*,\, \boldsymbol{\Sigma}^*,\, K)} \times \\ &\quad [b(u_1;2,2)\,b(u_2;2,2)\,b(w_{11};3,3)\,b(w_{12};3,3)\,b(w_{22};3,3)]^{-1}\,. \end{aligned}$$

The first term in R_{prop} is the ratio $\Pr(R = m \mid K+1)/\Pr(R = s \mid K)$. The second term is the ratio of selecting two classes to merge (when there are $K+1$) classes to that of selecting a class to split (when there are K classes). The third term is the ratio of allocation probabilities. The fourth term accounts for the random draws of the five new parameters necessary for dimension matching where $b(x;\alpha,\beta)$ is the density of a beta random variable, x, with parameters α and β.

R_{prior} is the prior ratio, given by

$$\begin{aligned} R_{\text{prior}} &= \frac{\lambda}{K+1} \times [(\max \mathbf{Y}_F - \min \mathbf{Y}_F)(\max \mathbf{Y}_R - \min \mathbf{Y}_R)]^{-1} \times \\ & \quad w^{-1}(\Sigma_{k^*}, \nu_1,\, S)\, w^{-1}(\Sigma_{k^{**}}, \nu_1,\, S)\, w^{-1}(\Sigma_k, \nu_1,\, S) \times \\ & \quad \frac{\exp\left\{-\beta \sum_{i\sim j}[1-\delta_j(\mathcal{E}^*_{Ri})]\right\} \Big/ g(\beta,\, K+1)}{\exp\left\{-\beta \sum_{i\sim j}[1-\delta_j(\mathcal{E}_{Ri})]\right\} \Big/ g(\beta,\, K)}. \end{aligned}$$

The first term in R_{prior} is the prior ratio of the number of classes. The second term is the prior ratio of the means and the third term is the prior ratio of the covariance matrices. We note that $w^{-1}(\Psi; \nu,\, S)$ is the inverse Wishart density at Ψ with ν degrees of freedom and scale matrix S. The last term is the prior ratio of the class label maps.

R_{likeli} is the likelihood ratio, given by

$$R_{\text{likeli}} = \prod_{i=1}^{N} \frac{\left|\Sigma^{-1}_{Z^*_i}\right| \exp[-0.5(\mathbf{y}_i - \boldsymbol{\mu}_{z^*_i})^{\mathrm{T}} \Sigma^{-1}_{z^*_i} (\mathbf{y}_i - \boldsymbol{\mu}_{z^*_i})]}{\left|\Sigma^{-1}_{Z_i}\right| \exp[-0.5(\mathbf{y}_i - \boldsymbol{\mu}_{z_i})^{\mathrm{T}} \Sigma^{-1}_{z_i} (\mathbf{y}_i - \boldsymbol{\mu}_{z_i})]}.$$

The acceptance probability for a merge proposal, assuming there are currently K classes, is given by $\min(1, A_s^{-1})$ with K and $K+1$ replaced by $K-1$ and K, respectively, in R_{prior} and in the first two terms of R_{prop}.

References

Besag, J. E. (1974). Spatial interaction and the statistical analysis of lattice systems. *Journal of the Royal Statistical Society, Series B* **36**, 192–236.

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

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

Green, P. J. and Richardson, S. (2002). Hidden markov models and disease mapping. *Journal of the American Statistical Association* **97**, 1055–1070.

Halmos, P. R. (1998). *Naive Set Theory*. 1st edition. New York: Springer.

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

Haubner, R., et al. (1999). Radiolabeled $\alpha_v\beta_3$ integrin antagonists: a new class of tracers for tumor targeting. *Journal of Nuclear Medicine* **40**, 1061–1071.

Higdon, D. M. (1998). Auxiliary variable methods for Markov chain Monte Carlo with applications. *Journal of the American Statistical Association* **93**, 585–595.

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

Piert, M., et al. (2005). Hypoxia-specific tumor imaging with ^{18}F-Fluoroazomycin Arabinoside. *Journal of Nuclear Medicine* **46**, 106–113.

Potts, R. B. (1952). Some generalized order-disorder transformations. *Proceedings of the Cambridge Philosophic Society* **48**, 106–109.

Richardson, S. and Green, P. J. (1997). On Bayseian analysis of mixtures with an unknown number of components (with discussion). *J. R. Statist. Soc.* ***B*** **59**, 731–792.

Robert, C. P. (2001). *The Bayesian Choice*. 2nd edition. New York: Springer.

Robert, C. P. and Casella, G. (2005). *Monte Carlo Statistical Methods*. 2nd edition. New York: Springer.

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

Table 1: Bias results from simulation study. Entries in the table are the average bias from the 20 simulations for 4 different models. The true parameter values for the eight classes are listed at the bottom of the table.

Parm.	Model	Class 1	2	3	4	5	6	7	8
μ_F	Spat, Bivar	-0.002	0.010	0.009	0.006	-0.003	0.009	0.018	-0.006
	Spat, Bivar, K = 4	-0.274	-0.135	0.293	0.154	-0.122	0.733	0.198	-0.810
	Spat, Uncorr	0.148	-0.133	-0.124	0.143	-0.718	0.716	0.717	-0.718
	Non Spat, Bivar	-0.222	0.263	0.244	-0.243	-0.511	0.527	0.511	-0.529
μ_R	Spat, Bivar	-0.009	-0.003	0.002	0.020	-0.009	-0.006	0.006	0.007
	Spat, Bivar, K = 4	-0.421	0.215	0.118	-0.149	-0.056	-0.641	0.130	0.811
	Spat, Uncorr	0.125	0.124	-0.127	-0.120	-0.713	-0.712	0.716	0.713
	Non Spat, Bivar	-0.240	-0.243	0.236	0.253	-0.505	-0.522	0.501	0.528
σ_F	Spat, Bivar	-0.042	-0.039	-0.042	-0.041	-0.014	-0.027	-0.021	-0.006
	Spat, Bivar, K = 4	0.140	0.157	0.142	0.158	0.045	0.146	0.042	0.162
	Spat, Uncorr	0.503	0.478	0.478	0.503	0.506	0.475	0.475	0.506
	Non Spat, Bivar	-0.040	-0.043	-0.045	-0.042	0.006	-0.011	-0.020	-0.003
σ_R	Spat, Bivar	-0.049	-0.040	-0.042	-0.039	-0.009	-0.027	-0.018	-0.008
	Spat, Bivar, K = 4	0.142	0.129	0.113	0.148	0.011	0.118	0.022	0.151
	Spat, Uncorr	0.477	0.477	0.489	0.489	0.475	0.475	0.488	0.487
	Non Spat, Bivar	-0.048	-0.049	-0.045	-0.049	-0.009	-0.015	-0.019	0.000
ρ	Spat, Bivar	0.020	-0.019	0.022	-0.021	0.000	-0.011	0.012	-0.003
	Spat, Bivar, K = 4	0.647	-0.642	0.586	-0.705	0.069	-0.565	0.126	-0.677
	Spat, Uncorr	0.750	-0.750	0.750	-0.750	0.750	-0.750	0.750	-0.750
	Non Spat, Bivar	0.817	-0.843	0.797	-0.861	0.437	-0.449	0.433	-0.464

True Parameter Values

Class	μ_F	μ_R	σ_F	σ_R	ρ	Class	μ_F	μ_R	σ_F	σ_R	ρ
1	0.5	0.5	1.0	1.0	-0.75	2	1.5	1.5	1.0	1.0	-0.75
3	-0.5	0.5	1.0	1.0	0.75	4	-1.5	1.5	1.0	1.0	0.75
5	-0.5	-0.5	1.0	1.0	-0.75	6	-1.5	-1.5	1.0	1.0	-0.75
7	0.5	-0.5	1.0	1.0	0.75	8	1.5	-1.5	1.0	1.0	0.75

Table 2: The mean misclassification and odds ratio of misclassification for the 20 simulated data sets. The odds ratio of misclassification is defined as the ratio of the odds of misclassification for the last 4 models to the odds of our model, model 1.

	Mean	
Model	Misclass.	O.R.
Spatial, Bivariate	0.25%	-
Spatial, Uncorrelated	6.43%	25.33
Spatial, Bivariated, K = 4	29.29%	212.76
Empirical, data	39.17%	269.71
Non-Spatial, Bivariate	42.97%	316.32

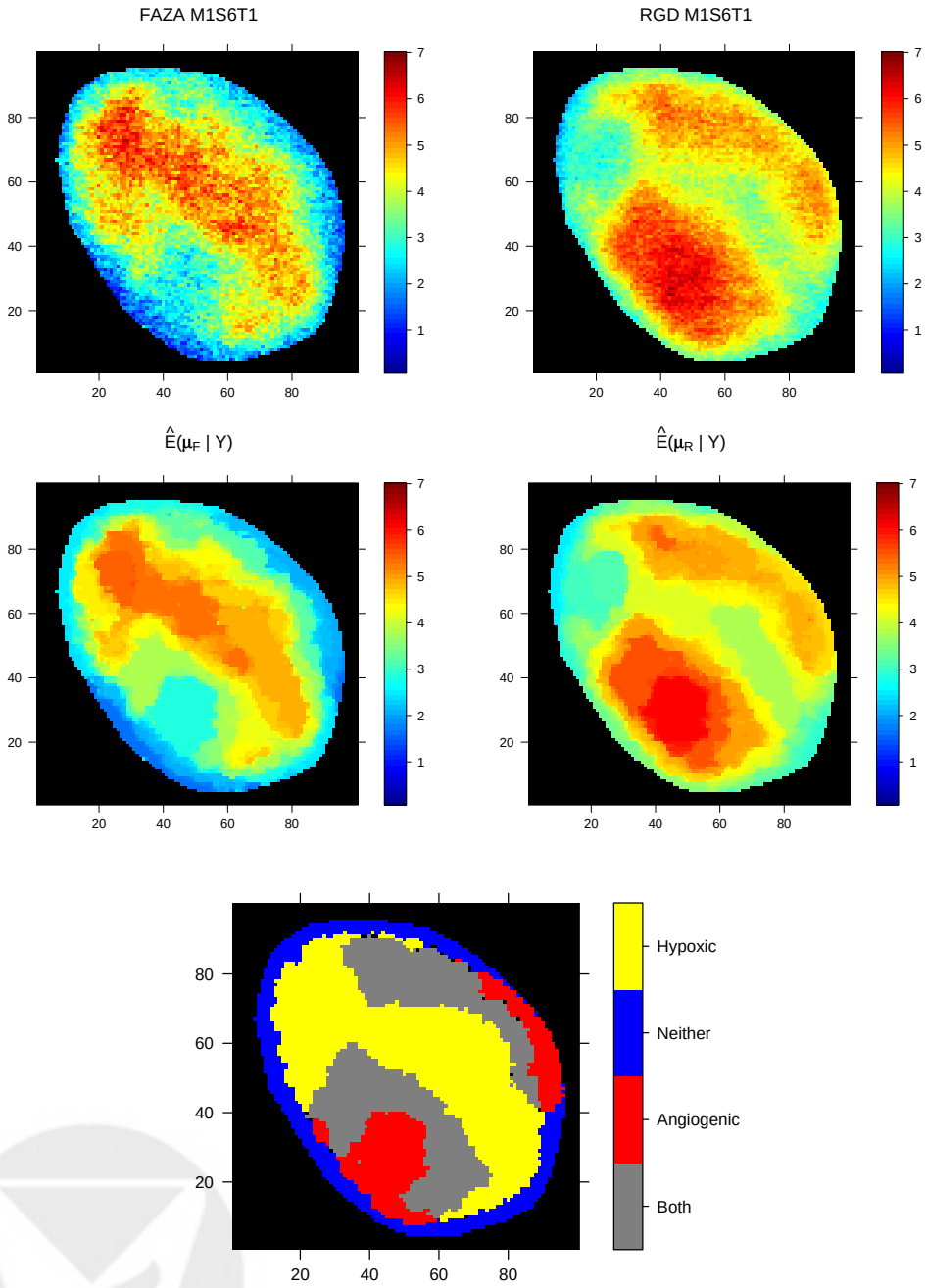

Figure 1: Slice M1S6T1. The top row shows the FAZA and RGD observed intensities. The posterior mean intensity images are displayed in the middle row. The bottom image displays the subregions of the tumor where both FAZA and RGD intensity are high (gray), where both are low (blue), where FAZA is high (yellow) and where RGD is high (red).

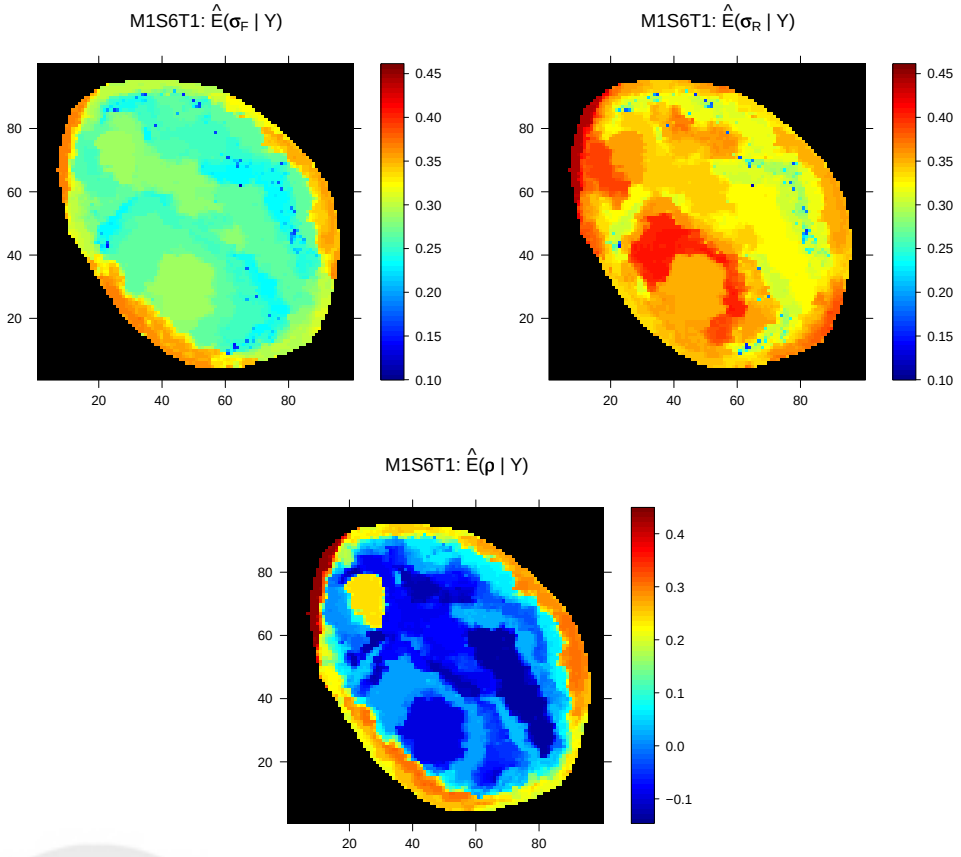

Figure 2: Posterior estimates of the standard deviations (slice M1S6T1) of the tracers as well as the correlation between them.

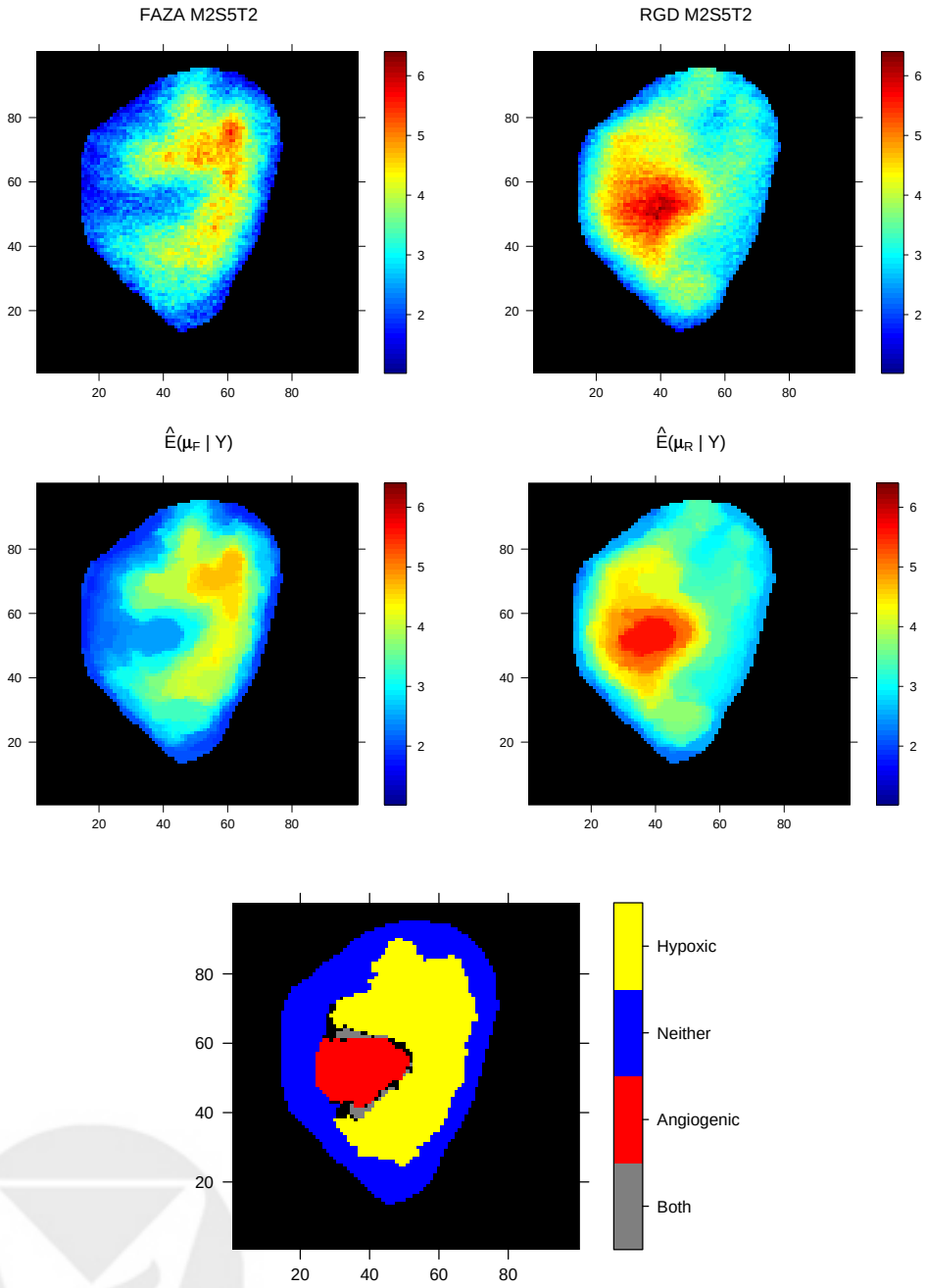

Figure 3: Slice M2S5T2. The top row shows the FAZA and RGD observed intensities. The posterior mean intensity images are displayed in the middle row. The bottom image displays the subregions of the tumor where both FAZA and RGD intensity are high (gray), where both are low (blue), where FAZA is high (yellow) and where RGD is high (red).

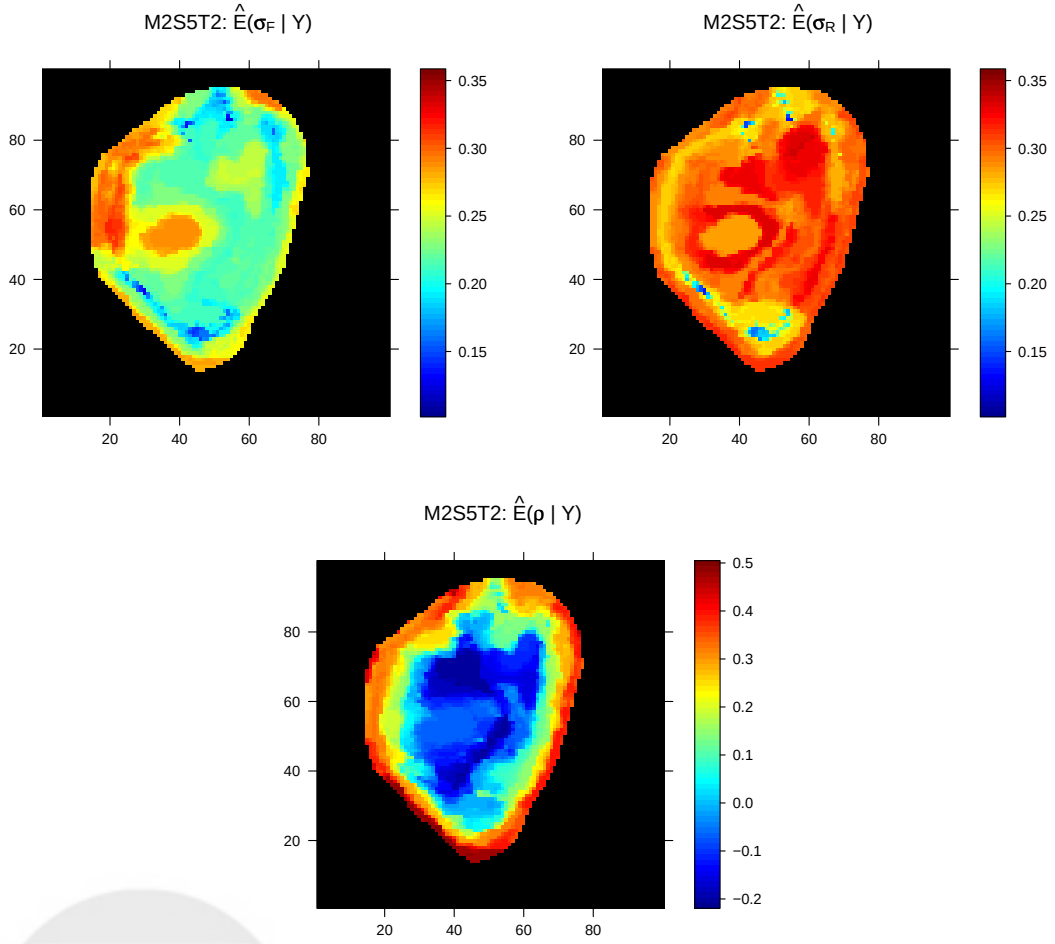

Figure 4: Posterior estimates of the standard deviations (slice M2S5T2) of the tracers as well as the correlation between them.

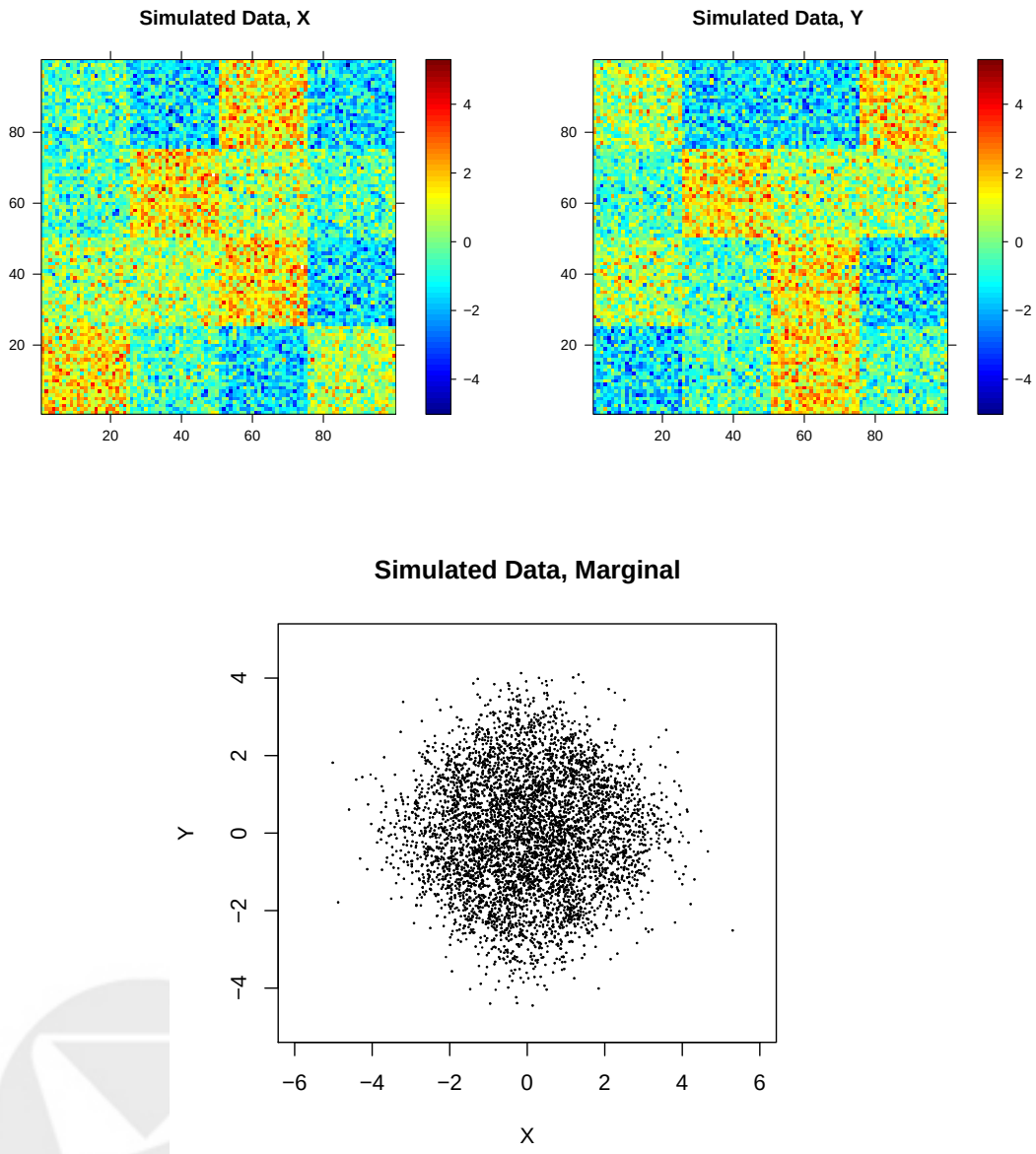

Figure 5: An example of one of the simulated data sets. The top row shows the simulated images. The bottom row displays the marginal (ignoring the spatial structure) data.

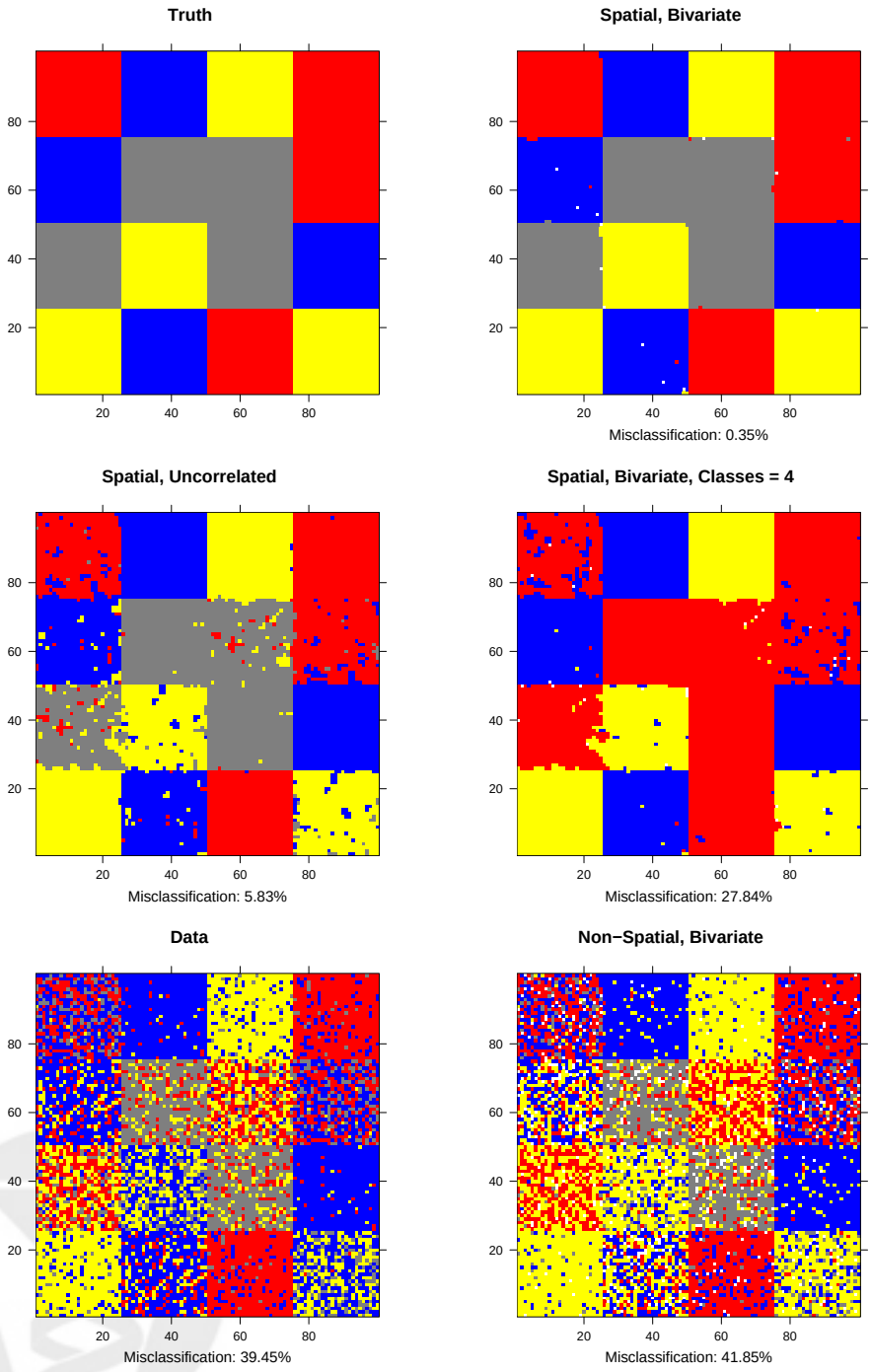

Figure 6: Classification results for the simulated data shown in Figure 5. Clearly, our proposed model outperforms the others in this example. Note that the spatial bivariate model with $K = 4$ completely misses on of the classes (gray). Gray: Both higher than 0.4. Blue: Neither higher than 0.4. Yellow: X higher, Y lower. Red: X lower, Y higher.

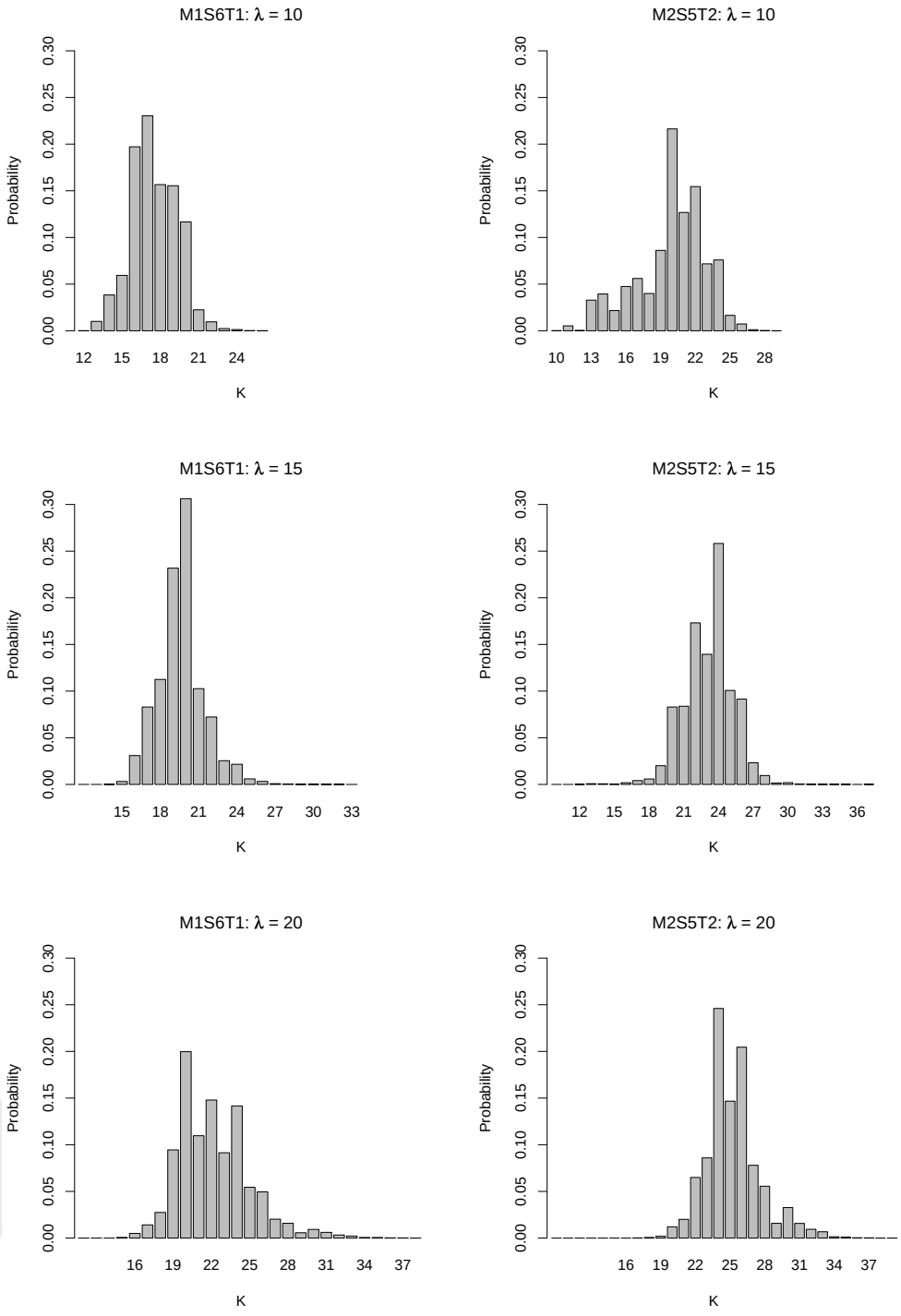

Figure 7: Posterior estimates of the number of equivalence classes. As the prior parameter λ increases from 10 to 15 to 20, the distributions shift to the right.