

MRI Tissue Classification Using High Resolution
Bayesian Hidden Markov Normal Mixture Models

Dai Feng

Biometrics Research Department

Merck Research Laboratories

Rahway, NJ 07065

Luke Tierney

Department of Statistics & Actuarial Science

University of Iowa

Iowa City, IA 52242

September 2, 2009

Dai Feng is Biometrician, Biometrics Research Department, Merck Research Laboratories, RY-33 300, P.O.Box 2000, Rahway, NJ 07065 (E-mail: dai_feng@merck.com). Luke Tierney is Professor, Department of Statistics & Actuarial Science, University of Iowa, Iowa City, IA 52242 (E-mail: luke@stat.uiowa.edu). Research was supported in part by National Science Foundation grant DMS 06-04593. Some of the computations for this paper were performed on equipment funded by National Science Foundation grant DMS 06-18883. The authors would like to thank Dr. Vincent Magnotta for helpful discussions and for making available the Iowa Mental Health Clinical Research Center data.

Abstract

Magnetic resonance imaging (MRI) is used to identify the major tissues within a subject's brain. Classification is usually based on a single image providing one measurement for each volume element, or voxel, in a discretization of the brain. A simple model views each voxel as homogeneous, belonging entirely to one of the three major tissue types (gray matter, white matter, and cerebro-spinal fluid); the measurements are normally distributed with means and variances depending on the tissue types of their voxels. Since nearby voxels tend to be of the same tissue type, a Markov random field model can be used to capture the spatial similarity of voxels. A more realistic model would take into account the fact that some voxels are not homogeneous and contain tissues of more than one type. Our approach to this problem is to construct a higher resolution image in which each voxel is divided into subvoxels, and subvoxels are in turn assumed to be homogeneous and follow a Markov random field model. This paper uses a Bayesian hierarchical model to conduct MRI tissue classification. Conditional independence is exploited to improve the speed of sampling. The subvoxel approach provides more accurate tissue classification and also allows more effective estimation of the proportion of major tissue types present in each voxel for both simulated and real data sets.

KEY WORDS: Markov random field; Conditional independence; Markov chain Monte Carlo; Brain imaging; Partial volume effect;

1 INTRODUCTION

Magnetic resonance imaging (MRI) is an important non-invasive method for visualizing the human brain and for examining the brain’s anatomical structure. An MR image is based on a discretization of the viewing area into a three-dimensional (3D) array of volume elements, or voxels. An image provides one measurement, the intensity value, for each voxel. For a single subject (the brain), there could be different types of images, based on multiple measurements for each voxel. MRI tissue classification (segmentation) is a process during which anatomically meaningful labels are assigned to each voxel of the image. Gray matter (GM), white matter (WM), and cerebrospinal fluid (CSF) are the three main tissue types of the brain. There are others, but tissue classification usually focuses on these three. MRI segmentation allows the 3D visualization of individual anatomical structures, which facilitates the study of abnormal tissue, diagnosis of disease, and surgery and treatment planning. After classification the volumes of certain tissues can be calculated, thereby providing important quantitative information for the study of diseases such as Alzheimer’s disease, schizophrenia, and multiple sclerosis.

Since manual tissue classification is very labor intensive, researchers have been trying to develop automated methods that match the quality of manual classification at lower cost. Usually the classification is based on intensities in a particular type of MR image called a T1 image. In a T1 MR image (see Figure 1 for example), WM has high intensities (light gray), GM has medium intensities (medium gray), and CSF has low intensities (dark gray).

MR image data consist of image intensities $\mathbf{y} = (y_1, \dots, y_N)$ for N voxels in a 3D grid. For the examples used in this paper, the value of N is either $256 \times 256 \times 192$ or $181 \times 217 \times 181$. To allow storage as a single byte, intensities are often scaled to the range $[0, 255]$ and rounded to an integer. A simple model views each voxel as homogeneous. The problem is then to determine the tissue type $z_i \in \{1, \dots, k\}$ for voxel i , where $i = 1, 2, \dots, N$, and $k = 3$

corresponding to three major tissue types. A more realistic model takes into account the fact that some voxels are not homogeneous and contain more than one type of tissue; this is sometimes referred to in the literature as the partial volume effect. Since adjacent voxels tend to contain the same tissue types a spatial model, such as a Markov random field, that captures this similarity can improve the classification.

The next section reviews some existing approaches to MRI tissue classification and outlines a new approach for handling the partial volume effect. Section 3 presents the Markov random field models used to capture spatial similarity of adjacent voxels, describes the prior and full conditional posterior distributions of other parameters, and discusses related computational issues. The results of classifications performed on simulated and real data are shown in Section 4. Some discussion and directions for future work are presented in Section 5.

2 SOME EXISTING METHODS AND A NEW APPROACH

A variety of statistical and other methods have been used in MRI tissue classification; see Archibald et al. (2003), Cocosco et al. (2003), Cuadra et al. (2005), Harris et al. (1999), Kamber et al. (1995), Leemput et al. (1999), Warfield et al. (2000), Shattuck et al. (2001) and references therein. This section outlines some approaches based on a pure voxel assumption, some existing approaches to the partial volume issue, and a new approach based on a higher resolution model.

2.1 Methods Based on the Pure Voxel Assumption

Density plots of the intensities in MR images often exhibit multi-modality. A density plot of the intensities in a relatively low noise MR image from Collins et al. (1998) is shown in Figure 2. See Section 4 for details of the data.

A mixture model is a natural choice when there are multiple modes in a density plot. A simple model views each voxel as homogeneous, belonging entirely to one of the three major tissue types, and, given the tissue structure $\mathbf{z} = \{z_1, \dots, z_N\}$, the index vector of tissue types of all voxels, intensities are independently and normally distributed with

$$y_i | z_i \sim N(\mu(z_i), \sigma^2(z_i))$$

where means and variances depend on the tissue type. If we assume tissue types are independent, then this leads to a simple normal mixture model

$$f(\mathbf{y}) = \prod_{i=1}^N \sum_{j=1}^k \phi_{\mu(z_i), 1/\sigma^2(z_i)}(y_i) p(z_i = j) \quad (1)$$

Throughout the paper $\phi_{\mu, \tau}$ denotes the probability density function of a normal distribution with mean μ and precision τ . The parameters of model (1) are easily estimated by the EM algorithm, and tissue types can be assigned using the Bayes classifier. The normal mixture model is the basic setup in MRI tissue classification upon which various methods taking other factors into consideration are built. See Cuadra et al. (2005) and references therein.

Since the data are collected on a 3D grid, a natural question is how to incorporate spatial information to improve the classification. Since adjacent voxels are likely to contain tissues of the same type a more realistic model would account for spatial homogeneity in \mathbf{z} . The Potts model family, introduced in more detail in Section 3, provides simple models for assigning spatial homogeneity relationships among the components of \mathbf{z} . After incorporating

spatial information through Potts models, which belong to the family of Markov random field models, we have the hidden Markov normal mixture model

$$p(\mathbf{y}, \mathbf{z} | \boldsymbol{\mu}, \boldsymbol{\tau}) = p(\mathbf{y} | \mathbf{z}, \boldsymbol{\mu}, \boldsymbol{\tau}) p(\mathbf{z}) \quad (2)$$

The normal mixture model is used to model the intensities and, given the tissue types, all intensities are independently normally distributed. The spatial homogeneity relationship among voxels is modeled by using a hidden Markov model on the component index vector of the normal mixture model.

The iterated conditional mode (ICM) algorithm of Besag (1986) can be used to estimate parameters of the hidden Markov normal mixture model. For the ICM approach we fit the model (2) by alternately maximizing over $\boldsymbol{\mu}, \boldsymbol{\tau}$ for fixed \mathbf{z} and maximizing over each z_i for fixed values of $\boldsymbol{\mu}, \boldsymbol{\tau}$ and the other \mathbf{z} values. Though the ICM has its drawbacks, such as sensitivity to the initial configuration and visiting scheme, and can get stuck at a local maximum, it is fast and easy to implement. For a detailed discussion of ICM see Winkler (2003).

The hidden Markov random field EM (HMRF-EM) algorithm proposed in Zhang et al. (2001) is another algorithm that can be used to fit the hidden Markov normal mixture model. HMRF-EM is also an iterative algorithm that alternates estimating \mathbf{z} for given $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$ with several iterations of the ICM algorithm and then, for a given \mathbf{z} , estimates $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$ at iteration $t + 1$ by

$$\begin{aligned} \mu_j^{(t+1)} &= \frac{\sum_{i=1}^N p(z_i^{(t)} = j | y_i, \boldsymbol{\mu}^{(t)}, \boldsymbol{\tau}^{(t)}, z_{\partial(i)}^{(t)}) y_i}{\sum_{i=1}^N p(z_i^{(t)} = j | y_i, \boldsymbol{\mu}^{(t)}, \boldsymbol{\tau}^{(t)}, z_{\partial(i)}^{(t)})} \\ \frac{1}{\tau_j^{(t+1)}} &= \frac{\sum_{i=1}^N p(z_i^{(t)} = j | y_i, \boldsymbol{\mu}^{(t)}, \boldsymbol{\tau}^{(t)}, z_{\partial(i)}^{(t)}) (y_i - \mu_j^{(t+1)})^2}{\sum_{i=1}^N p(z_i^{(t)} = j | y_i, \boldsymbol{\mu}^{(t)}, \boldsymbol{\tau}^{(t)}, z_{\partial(i)}^{(t)})} \end{aligned}$$

where $p(z_i^{(t)} = j | y_i, \boldsymbol{\mu}^{(t)}, \boldsymbol{\tau}^{(t)}, z_{\partial(i)}^{(t)})$ is the conditional probability that $z_i = j$ given the voxel intensity y_i , the current values of $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$, and the tissue types of the neighbors of voxel i , $z_{\partial(i)}$ at iteration t . This process is continued until a convergence criterion is met.

Instead of using ICM or HMRF-EM, a Bayesian approach could be used to study the hidden Markov normal mixture model. The model is formulated in a hierarchical framework. Given the tissue types, the intensities are independently and normally distributed. A model from the Potts model family is used as a prior distribution of the tissue type indices. Conjugate prior distributions are assigned on the means and variances of the normal mixture model. In image analysis, mixture models with spatial Markovian dependency studied in a Bayesian framework have been used, for example, in animal tissue classification (Alston et al., 2005) and fMRI data segmentation (Woolrich et al., 2005).

2.2 Methods Used to Address the Partial Volume Effect

The approaches described in Section 2.1s assume that each voxels contains only one tissue type. A more realistic model recognizes that a voxel may contain one, two, or possibly three different tissue types. This phenomenon is called the partial volume effect (or PV effect).

An approach to addressing the PV effect used by several studies is to introduce two additional classes into consideration besides CSF, GM, and WM. The two additional classes are the combination of CSF and GM (CG) and the combination of GM and WM (GW). Voxels containing WM and CSF are very rare and are ignored; this helps to reduce confounding in estimation. Therefore, instead of three, there are five major tissue types: CSF, CG, GM, GW, and WM, and the goal is to determine $z_i \in \{1, 2, 3, 4, 5\}$ corresponding to five tissue types for each voxel $i = 1, \dots, N$.

First, a simple normal mixture model with five components, one component corresponding to one tissue type, could be used. Second, a hidden Markov normal mixture model with five components could be a better choice, since it takes spatial information into account.

The estimation of the two models could be conducted by using the ICM or the modified EM algorithms.

Instead of allowing the five mean and variance parameters of the mixture components to vary independently Santago and Gage (1993) and Santago and Gage (1995) propose a third approach in which the means and variances for the combined categories are related to the means and variances for the pure categories by

$$\begin{aligned}\mu_2 &= \alpha\mu_1 + (1 - \alpha)\mu_3 \\ \mu_4 &= \alpha\mu_3 + (1 - \alpha)\mu_5\end{aligned}$$

and

$$\begin{aligned}\sigma_2^2 &= \alpha^2\sigma_1^2 + (1 - \alpha)^2\sigma_3^2 \\ \sigma_4^2 &= \alpha^2\sigma_3^2 + (1 - \alpha)^2\sigma_5^2\end{aligned}$$

with $0 < \alpha < 1$ and α is assumed to follow some mixing distribution $p(\alpha)$. When $z_i \in \{2, 4\}$

$$p(y_i|z_i) = \int_0^1 \phi_{\mu_{z_i}, 1/\sigma_{z_i}^2}(y_i)p(\alpha|z_i)d\alpha \quad (3)$$

Choosing the distribution of α is a complex issue and it is often assumed to be uniform on $[0, 1]$. Then equation (3) reduces to

$$p(y_i|z_i) = \int_0^1 \phi_{\mu_{z_i}, \sigma_{z_i}^2}(y_i)d\alpha \quad (4)$$

There is no closed form solution to the integral and it has to be computed numerically. The model can be fitted by minimizing of the sum squares of the differences between the observed normalized intensity densities and the densities from the model using a genetic algorithm

(Schroeter et al., 1998).

The fourth model is the third model plus a Potts prior on \mathbf{z} and can be fitted by the modified EM algorithm. This model is referred to as GPV-HMRF hereafter.

Besides the four parametric approaches, some non-parametric methods are discussed in Cuadra et al. (2005) as well. Compared with GPV-HMRF, the other methods appear to be less competitive in performance. Therefore we do not consider them further.

2.3 A High Resolution Model

Instead of addressing the PV effect at the voxel level, suppose each voxel is made up of n_{sub} subvoxels and what we observe is the intensity of the whole voxel denoted by y . For example, if each voxel is divided in half in the x , y , and z directions then this produces $n_{sub} = 8$ subvoxels. This is illustrated in Figure 3. Let $v_i, i = 1, 2, \dots, n_{sub}$, be the unobserved subvoxel intensities for a given voxel with observed intensity y . We then assume that the observed intensity is the sum of the corresponding unobserved subvoxel intensities, $y = \sum_{i=1}^{n_{sub}} v_i$. The subvoxels are in turn assumed to be homogeneous and follow the hidden Markov normal mixture model based on the pure voxel assumption described in Section 2.1. In the Markov random field model it is important to take into account the observation that WM and CSF are rarely adjacent. This is discussed further in Section 3.

3 PRIOR DISTRIBUTIONS AND POSTERIOR SAMPLING

The joint distribution of the observed voxel intensities \mathbf{y} , the unobserved subvoxel intensities \mathbf{v} , the normal distribution parameters $\boldsymbol{\mu}$ and $\boldsymbol{\tau}$, and the subvoxel tissue membership

indicators \mathbf{z} is assumed to be of the form

$$p(\mathbf{y}|\mathbf{v})p(\mathbf{v}|\boldsymbol{\mu}, \boldsymbol{\tau}, \mathbf{z})p(\boldsymbol{\mu})p(\boldsymbol{\tau})p(\mathbf{z}) \quad (5)$$

Posterior sampling uses an MCMC algorithm that cycles through sampling from the full conditional distributions for \mathbf{z} , \mathbf{v} , $\boldsymbol{\mu}$, and $\boldsymbol{\tau}$. This section describes the components of the prior distributions, the resulting full conditional distributions, and the sampling approaches used.

3.1 Prior and Full Conditional Posterior Distributions of Tissue Types.

A Markov random field model is used to capture spatial similarity of adjacent voxels. We use a six-neighbor structure in which voxels or subvoxels are viewed as neighbors if they share a face. These are also sometimes referred to as first order neighbors. A commonly used model is of the form

$$p(\mathbf{z}) = C(\beta)^{-1} \exp \left\{ \sum_{i=1}^N \alpha_i(z_i) + \beta \sum_{i \sim j} w_{ij} f(z_i, z_j) \right\} \quad (6)$$

where $C(\beta)$ is a normalizing constant and $i \sim j$ indicates neighboring vertices. The non-negative parameter β , called inverse temperature, determines the level of spatial homogeneity between neighboring vertices in the image. A value of $\beta = 0$ would imply that neighboring vertices are independent. The w_{ij} are positive weights. The term $\sum_{i=1}^N \alpha_i(z_i)$ is called the *external field*. Prior distributions used to capture spatial similarity typically do not incorporate an external field; the corresponding full conditional posterior distribution based on independent observations on each voxel is then of the form of Equation (6) with an external field derived from the data likelihood.

The simplest form commonly used for $f(z_i, z_j)$ is $f(z_i, z_j) = \mathbf{I}(z_i = z_j)$. For $k = 2$

components the resulting model is called the Ising model (Ising, 1925); for $k > 2$ it is the Potts model (Potts, 1953). The Ising model was originally proposed to describe the physical properties of magnets. Due to its flexibility and simplicity, the Ising model and its variations have been widely used in other fields, such as brain models in cognitive science, information and machine learning theory (see for example MacKay, 2003), economics (see for example Bourguine and Nadal, 2004), sociology (Kohring, 1996) and game theory (Hauert and Szabó, 2005).

Another useful form of the function $f(z_i, z_j)$ is $f(z_i, z_j)$ as

$$f(z_i, z_j) = \begin{cases} a_1 & \text{if } z_i = z_j \\ a_2 & \text{if } |z_i - z_j| = 1 \\ a_3 & \text{otherwise} \end{cases} \quad (7)$$

where $a_1 \geq a_2 \geq a_3$. A special case of this model is used in Cuadra et al. (2005). We use this model with $\beta a_1 = \beta_1$, $a_2 = 0$, and $\beta a_3 = -\beta_2$, with $\beta_1, \beta_2 > 0$, which leads to

$$p(\mathbf{z}) = C(\beta_1, \beta_2)^{-1} \exp \left\{ \sum_{i \sim j} \beta_1 w_{ij} \mathbf{I}(z_i = z_j) - \beta_2 w_{ij} \mathbf{I}(|z_i - z_j| > 1) \right\} \quad (8)$$

We call this model the *repulsion Potts model* and β_2 the *repulsion parameter*. This model assumes an ordering of the categories and that neighboring vertices are most likely from the same category, and, if they are different then it is more likely that they are similar than totally different. For brain tissue classification we associate the index values 1, 2, 3 with categories CSF, GM, and WM, respectively, and use the repulsion Potts model to reflect the observation that CSF and WM are rarely adjacent. A simple Potts model that does not account for this cannot effectively distinguish pure GM voxels from voxels combining WM and CSF, which are in fact very rare.

We follow common practice in the imaging literature (e.g. Cuadra et al., 2005) and determine the parameters of the Markov random field prior on \mathbf{z} by a pilot experiment; more details are given in Section 4. As a result, the normalizing constant in (6) or (8) is not needed. An alternative would be to place a prior distribution on the β parameters and to use some approaches to computing or approximating the normalizing constant. Gelman and Meng (1998) gave a thorough discussion on using simulation to compute normalizing constants. We could use the thermodynamic integration approach, whose name comes from the differential equations for describing thermodynamic relationships in physics, to obtain the normalizing constants. This method was used in Green and Richardson (2002) and Smith and Smith (2006). They used the simple Potts model or Ising model defined on a two dimensional graph. A method incorporating the techniques of kernel density estimation into simulations was proposed in Liang (2006) to compute the normalizing constants. In Møller et al. (2006), the ratio of the normalizing constants is canceled out by the introduction of an auxiliary variable in a Metropolis-Hastings algorithm.

For the examples considered in this paper and for one modality (T1,T2 or PD) for each subject, after masking out some points outside the brain there remain nearly two million observations. It is not quite clear whether existing methods can find the normalising constants of a Potts model without external field defined on a vast space in MRI tissue classification accurately with reasonable costs. It is also not clear whether the classification accuracy of the approach would be improved by these efforts.

To generate random samples from the full conditional distribution of \mathbf{z} , which is of the form of Equation (6) with an external field, we use a single site updating Gibbs sampler but with an updating order that takes advantage of the conditional independence structure to update index variables $z_i, i = 1, 2, \dots, N$. This allows the simulation to be vectorized in high level languages like R or Matlab and also can form the basis for a parallel computing strategy.

The approach is illustrated in Figure 4 (a), for a four neighbor structure in 2D, and Figure 4 (b) for a six neighbor structure in 3D. Given the categories of the black cells, the categories of the white cells are independent and vice versa. As a result, the vector \mathbf{z} can be updated in two steps: one for the black cells, one for the white cells. This idea was described in Wilkinson (2005) and detailed discussion can be found in Winkler (2003). The idea of taking advantage of this kind of independence can be traced back at least to the ‘‘Coding Methods’’ in Besag (1974). The minimum number of blocks to make the vertices within each block independent given the other blocks is called the *chromatic number* in Winkler (2003). The chromatic numbers for a four neighbor configuration in 2D and six neighbors in 3D are both 2. Under the eight neighbor structure in 2D, the eighteen neighbor structure in 3D, and the twenty-six neighbor structure in 3D, the chromatic numbers are 4, 7, and 8, respectively (Feng, 2008). Our current implementation in C uses Open MP (Chandra et al., 2000) to parallelize loops in the generation within each block and achieves a speedup of roughly a factor of two on a four processor machine; a more sophisticated approach should be able to realize further gains.

The full conditional distribution of the indicator variable z_i for subvoxel i is

$$\begin{aligned} p(z_i | \mathbf{z}_{\partial(i)}, \boldsymbol{\mu}, \boldsymbol{\tau}, \mathbf{v}) &\propto p(v_i | z_i, \boldsymbol{\mu}, \boldsymbol{\tau}) p(z_i | \mathbf{z}_{\partial(i)}) \\ &\propto \text{pdf of multinomial}(1; p_{i1}, p_{i2}, p_{i3}) \end{aligned}$$

where $\mathbf{z}_{\partial(z_i)}$ represents the indices for components of all neighbors of subvoxel i , and p_{ij} is the conditional probability that z_i is from component j . The p_{ij} are of the form

$$p_{ij} = \frac{\tau_j^{1/2} \exp \left\{ -\frac{\tau_j}{2} (v_i - \mu_j)^2 + r_{ij} \right\}}{\sum_{t=1}^3 \tau_t^{1/2} \exp \left\{ -\frac{\tau_t}{2} (v_i - \mu_t)^2 + r_{it} \right\}} \quad (9)$$

where r_{ij} determines the conditional prior probability that subvoxel i is from category j

given the categories of all neighbors of i . For model (8)

$$(r_{i1}, r_{i2}, r_{i3}) = (N_{i1}, N_{i2}, N_{i3}) \begin{pmatrix} \beta_1 & 0 & -\beta_2 \\ 0 & \beta_1 & 0 \\ -\beta_2 & 0 & \beta_1 \end{pmatrix}$$

where N_{ij} is the total number of neighbors of i which are currently allocated to category j .

For the simple Potts model, a number of alternatives to single site updating are available. The Swendsen and Wang (1987) algorithm (SW) is widely used to generate random samples from Potts models. Wolff's algorithm (Wolff, 1989) has been advocated as a better substitute to SW. Multigrid (Hurn, 1997, and references therein) and partial decoupling methods (Higdon, 1998) could be more powerful when dealing with models with external fields.

Various multiple site sampling methods might outperform single-site updating, but sometimes they might not. Multi-site sampling methods like the SW algorithm could tackle the critical slowing down problem. But when there is an external field, as in a posterior distribution like the one considered here, SW slows down since it does not make good use of the data. Hurn (1997) and Smith and Smith (2006) suggested that when β is large, Gibbs samplers might be more effective. For multigrid methods, questions such as how to decide which vertices should be considered simultaneously and make the Markov chains under the adaptive versions converge to the same target equilibrium distributions could be hard to solve. Furthermore there is no guarantee that this will be better than the Gibbs sampler as pointed out in Hurn (1997). The partial decoupling method might be better than SW, though it might take time to find the tuning parameters before running. For images as complicated as MRI, it is hard to determine the number of modes and even if we could, it might be the case that different tuning parameters facilitate transitions between different modes, not to mention the computational cost of trial and error. Besides, as pointed out in Hurn (1997), there are some vertices that cannot be updated simultaneously and sometimes

partial decoupling wouldn't be better than single-site updating.

For the simple Potts model, the best sampling method is likely to be problem-specific and there is no clear-cut winner as pointed out in Hurn (1997) and Higdon (1998). For the repulsion Potts model, possible multi-site updating methods are being investigated. Because of the strength of the external field in our application we believe that the mixing of the single site updating approach is likely to be competitive with other approaches, and the simplicity and opportunities for parallelization tip the balance in favor of this approach.

3.2 Updating Subvoxel Intensities

The full conditional distribution of the subvoxel intensities \mathbf{v} is a singular multivariate normal distribution in which the subvoxel values for different whole voxels are independent and the subvoxel intensities within a whole voxel sum to the observed voxel intensity. For an individual voxel with observed intensity y the $n_{\text{sub}} = 8$ subvoxel intensities $v_1, \dots, v_{n_{\text{sub}}}$ can be generated by the following algorithm:

Algorithm 3.1. *To generate a random draw from \mathbf{v} , with $\mathbf{v} \sim MVN(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, given $\sum_{i=1}^{n_{\text{sub}}} v_i = y$, where $\boldsymbol{\mu} = \{\mu_1, \dots, \mu_{n_{\text{sub}}}\}$ and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \dots, \sigma_{n_{\text{sub}}}^2)$:*

1. *Generate $\tilde{v}_i \stackrel{\text{ind}}{\sim} N(0, \sigma_i^2)$.*
2. *Compute $\tilde{\mathbf{v}}^* = \tilde{\mathbf{v}} - (\sum_{i=1}^{n_{\text{sub}}} \tilde{v}_i) \mathbf{a}$, where $\mathbf{a} = (a_1, a_2, \dots, a_{n_{\text{sub}}})$, $\tilde{\mathbf{v}} = (\tilde{v}_1, \tilde{v}_2, \dots, \tilde{v}_{n_{\text{sub}}})$, and*

$$a_i = \frac{\text{Cov}\left(v_i, \sum_{j=1}^{n_{\text{sub}}} v_j\right)}{\text{Var}\left(\sum_{j=1}^{n_{\text{sub}}} v_j\right)} = \frac{\sigma_i^2}{\sum_{j=1}^{n_{\text{sub}}} \sigma_j^2}$$

3. *Return $\tilde{\mathbf{v}}^* + \boldsymbol{\mu} + (y - \sum_{i=1}^{n_{\text{sub}}} \mu_i) \mathbf{a}$*

Proof. The conditional expectations are

$$E \left[v_i \left| \sum_{j=1}^{n_{\text{sub}}} v_j = y \right. \right] = \mu_i + a_i \left(y - \sum_{j=1}^{n_{\text{sub}}} \mu_j \right)$$

with a_i as defined above. In vector form this is

$$E \left[\mathbf{v} \left| \sum_{j=1}^{n_{\text{sub}}} v_j = y \right. \right] = \boldsymbol{\mu} + \left(y - \sum_{j=1}^{n_{\text{sub}}} \mu_j \right) \mathbf{a}$$

Now the residuals are

$$\mathbf{v}^* = \mathbf{v} - E \left[\mathbf{v} \left| \sum_{j=1}^{n_{\text{sub}}} v_j = y \right. \right] = \mathbf{v} - \boldsymbol{\mu} - \left(\sum_{j=1}^{n_{\text{sub}}} v_j - \sum_{j=1}^{n_{\text{sub}}} \mu_j \right) \mathbf{a}$$

So the conditional distribution $\mathbf{v} \mid \sum_{j=1}^{n_{\text{sub}}} v_j = y$ is the same as the distribution of

$$\mathbf{v}^* + \boldsymbol{\mu} + \left(y - \sum_{j=1}^{n_{\text{sub}}} \mu_j \right) \mathbf{a}$$

The distribution of \mathbf{v}^* does not depend on $\boldsymbol{\mu}$, so the distribution for general $\boldsymbol{\mu}$ is the same as the distribution for $\boldsymbol{\mu} = \mathbf{0}$, i.e. the same as the distribution of $\tilde{\mathbf{v}} - \mathbf{a} \sum_{i=1}^{n_{\text{sub}}} \tilde{v}_i$ with the $\tilde{v}_i \stackrel{\text{ind}}{\sim} \text{N}(0, \sigma_i^2)$. \square

The algorithm avoids matrix computations and can be easily vectorized.

3.3 Prior and Full Conditional Posterior Distributions of Means and Precisions

The prior distribution on (μ_1, μ_2, μ_3) has independent normals with the common mean ξ and common precision η conditioned to satisfy $\mu_1 < \mu_2 < \mu_3$. The ordering restriction avoids identifiability issues associated with “label switching” (Celeux et al., 2000). The

prior distributions of $\{\tau_j, j = 1, \dots, k\}$ are taken to be i.i.d. $\Gamma(\lambda, \phi)$ (the parametrization of the gamma distribution is with mean equal to $\lambda\phi$ and variance $\lambda\phi^2$). In our implementation the prior distribution parameters are chosen as follows. Let y_{\min} and y_{\max} be the minimum and maximum observed intensity values of voxels. Let $v_{\min} = \frac{y_{\min}}{n_{\text{sub}}}$ and $v_{\max} = \frac{y_{\max}}{n_{\text{sub}}}$. To specify ξ and η , we adopt the similar method as in Richardson and Green (1997). We let $\xi = \frac{v_{\min} + v_{\max}}{2}$ and let η be a small multiple m of $1/R^2$, with $R = v_{\max} - v_{\min}$. This produces a prior density that is fairly flat over the range supported by the data. For the prior distribution on the precisions let $\lambda = m$ and let $\phi = 1/R^2$. As to the value of m , the classification results are not very sensitive to it. We take m as 0.3, which works well across various data sets we tried. The full conditional distribution of $\boldsymbol{\mu}$ is

$$\mu_j \stackrel{\text{ind}}{\sim} \text{N} \left(\frac{\eta\xi + \tau_j \sum_{i \in C_j} v_i}{\eta + \tau_j N_j}, \eta + \tau_j N_j \right)$$

conditioned to satisfy $\mu_1 < \mu_2 < \mu_3$, where C_j is the set of all subvoxels currently allocated to component j and N_j is the size of C_j , i.e. the total number of subvoxels allocated to component j .

The full conditional distribution of $\boldsymbol{\tau}$ is

$$\tau_j \stackrel{\text{ind}}{\sim} \Gamma \left(\frac{N_j}{2} + \lambda, \frac{2\phi}{2 + \phi \sum_{i \in C_j} (v_i - \mu_j)^2} \right)$$

In our implementation initial values of $\{(\mu_j, \tau_j), j = 1, 2, 3\}$, are obtained by first running an EM algorithm for a simple mixture model at the whole voxel level (without taking spatial correlations into account) to obtain $\{(\mu_j^{\text{vox}}, \tau_j^{\text{vox}}), j = 1, 2, 3\}$, the means and precisions of each component of the normal mixture model on the distribution of the intensity values of voxels, and then computing initial subvoxel values as $\mu_j = \frac{\mu_j^{\text{vox}}}{n_{\text{sub}}}, \tau_j = n_{\text{sub}}^2 \tau_j^{\text{vox}}, j = 1, 2, 3$. The initial values of the components of \mathbf{z} are based on the initial values of $\{\mu_j, \tau_j, j = 1, \dots, k\}$, $z_i = m$, if $\max\{\phi_{\mu_j, \tau_j}(y_i), j = 1, \dots, k\} = m$.

4 APPLICATIONS

In this section we compare the subvoxel strategy proposed in Section 2.3 with alternative pure voxels approaches discussed in Section 2.1, and with methods addressing the PV effect discussed in Section 2.2. We used a simple Potts model with a six neighbor structure in 3D to capture the spatial homogeneity relationship among the components of the index vector of tissue types. Initial values of the means and precisions of the normal mixture model and the initial values of indices were chosen based on applying the EM algorithm to a simple normal mixture model with no spatial component analogously to the approach for the subvoxel method outlined in Section 3. All the β values for pure voxel methods are determined by pilot experiments as discussed later in this section. For the GPV-HMRF method, the β value is the the same as in Cuadra et al. (2005).

A challenge in assessing the performance of classification methods is to find data sets with a reliably known “ground truth.” We use both simulated and real data sets to evaluate the performance of the different methods.

4.1 Simulated Data Sets

The simulated data sets we used were from Collins et al. (1998) (referred to as the BrainWeb hereafter), which provides a Simulated Brain Database, a set of “true” MRI data volumes produced by an MRI simulator. This simulator uses an anatomical model and a model of MR acquisition physics to generate images with different nonuniformity (RF) and noise levels. The volume is $217 \times 181 \times 217$ voxels with isotropic 1 mm^3 voxel size. The anatomical model used to generate simulated images consists of a set of 3D “fuzzy” tissue membership volumes—the voxel values in these volumes reflect the proportion of each tissue present in that voxel. Besides the fuzzy tissue membership volumes, there is a discrete anatomical model which consists of a class label (integer) at each voxel, giving the tissue which con-

tributes the most to that voxel. See Kwan et al. (1999) for details of the simulator. Following Cuadra et al. (2005) and others, we consider RF values of 20% and 40% and noise levels between 1% and 9%. We use rfn_1pn_2 to represent data with RF n_1 percent and noise level n_2 percent. For example $rf20pn7$ represents the data with RF 20% and noise level 7%.

For the T1 images from the BrainWeb, some voxels in the images are not classified into CSF, GM, or WM: some are outside the brain and some are classified to other tissue types, such as glial matter. For real data, a brain mask is usually determined as a preliminary step and tissues other than CSF, GM, and WM are generally ignored. We therefore use a mask that restricts our attention to voxels containing only one of the three major tissue types or combinations of the three.

4.2 Determining β Values Using a Pilot Study

After obtaining data with the truth known, we first run pilot experiments to determine what β values to use for the different pure voxel methods. We split each image from the BrainWeb data into two halves, right and left, and use the right halves as training sets for determining good choices of the β values. For the simple Potts prior (used in the pure voxel models fitted by MCMC, ICM, and HMRF-EM), we consider the range $\beta \in (0, 4]$; for the repulsion Potts model, we let $\beta_1 \in (0, 4]$ and $\beta_2 \in (0, 4]$ also. For the repulsion Potts model used in the subvoxel method, we tried to find the best β values by grid searching on a coarser grid first and then on a more refined grid.

One way to measure the performance of different methods with different β s is to compute the *mis-classification rate*, the percentage of voxels not correctly classified (referred to as P_N hereafter). For the BrainWeb data, we assign the tissue type of each voxel using the discrete anatomical model giving the tissue which contributes the most to that voxel. For the Bayesian pure voxel model, the classification of a voxel i is based on the marginal posterior mode estimate of z_i . To allow the subvoxel method to be compared to whole voxel methods

we compute the posterior mean proportion of each tissue type within each whole voxel from the subvoxel results and then assign the entire voxel to the tissue with the highest posterior mean proportion. We will refer to this as the rounded subvoxel method.

We chose the β values as follows. For each method, first, let β_i be the best β (or (β_1, β_2) in the subvoxel method) for data set i , $i = 1, 2, \dots, T$, with T being equal to the number of data sets. We use $T = 10$ different data sets with various noise levels and nonuniformity settings. Figure 5 shows the settings used. Second, plug in each β_i into all data sets and calculate the largest mis-classification rate among all data sets P_N^i . Third, take the β which has the minimum P_N^i , $i = 1, 2, \dots, T$, as the best. The β values we obtained for each method are shown in Table 1. The P_N values are not very sensitive to small changes in the β values and we therefore expect that these values will work well for brain images with similar resolution, noise level, and nonuniformity.

For the Bayesian hidden Markov normal mixture model at the whole voxel level and for the subvoxel strategy, the number of sweeps used is 100, which takes about 1 and 16 minutes, respectively for the whole brain (run on a 2 dual core Opteron 2216, 2.4GHz, 8 GB RAM computer). This is of course a very short run for an MCMC approach, but it seems to be sufficient to produce good classifications. For example, for the rf20pn7 data, the results are comparable to results obtained from a run length of 20,000; the mis-classification rates change by only about 0.0002 for the whole voxel method and 0.002 for the subvoxel method. Trace plots for the means and variances of the normal mixture components based on runs of length 20,000 show no anomalies. We also randomly changed 10% of the initial values for the indices of tissue types and the chains converged to the same estimated images very quickly. Nevertheless, more systematic diagnostics would be useful and alternative sampling algorithms and implementations might improve the mixing and the performance of the samplers.

4.3 Empirical Results for Simulated and Real Data Sets

Having obtained β values using data from the right brain halves, we compare different methods for the simulated and real data sets. First we compare the subvoxel method with the other methods based on the assumption that all voxels are pure. Second, the comparison was done between the subvoxel method and the others addressing the PV effects.

4.3.1 Subvoxel Strategy and Pure Voxel Methods

To assess the performance of different methods, first we used the data from the left brain halves. Figure 5 shows the mis-classification rates for 10 data sets from the BrainWeb using the rounded subvoxel method and the other four pure voxel methods under consideration. From Figure 5 we can see that the simple normal mixture model fitted by the EM algorithm does not perform as well as the other methods, and that its performance deteriorates more rapidly as the noise level increases. The other methods perform similarly, and the larger the noise level, the greater the advantage of the Bayesian methods. The subvoxel strategy is the best when RF=20%, though by a small margin. Figure 6 shows the true classification and results of classification using the subvoxel method for a 2D sagittal slice from the left half of rf20pn7 data. From Figure 6, we can see that we have quite accurate classification results.

The mis-classification rates compare whole voxel classifications to the discrete anatomical model. This requires a form of rounding of the results and doesn't reflect all the partial volume information each approach provides. To illustrate the ability of the subvoxel approach to capture the partial volume effect, we computed average mean square errors between the tissue distributions within voxels estimated by the different methods and the "fuzzy" tissue membership information available for the BrainWeb data. The results are shown in Figure 7. Again the simple finite mixture model does not do as well as the other pure voxel approaches. As one might expect, the subvoxel approach does markedly better than the pure voxel methods.

In addition to the data from the BrainWeb, we also tested the classification methods on real data from the Iowa Mental Health Clinical Research Center (referred to as the Iowa data hereafter). For the Iowa data, we have 4 subjects and for each subject we have manually classified results for 2 slices. We use the Iowa data to evaluate the performance of classification methods in a setting where the truth is partially known. We use β values obtained from the BrainWeb data when analyzing the Iowa data. Figure 8 shows the misclassification rates of different methods. The subvoxel approach is uniformly best across all four subjects.

Figure 9 shows a slice of the T1 image of one subject, the manual classification results for the slice, and the results from the subvoxel classification. From Figure 9, we can see that the higher resolution images constructed by the subvoxel strategy are smoother and capture more detail. The black bits in the manually classified images are parts outside the brain, which cannot be captured by automatic methods unless they are masked out before the classification.

4.3.2 Subvoxel Strategy and PV Effect Methods

After comparing the subvoxel method with the four methods based on the three major tissue types classification, next we compare the subvoxel method with other methods addressing the PV effect discussed in Section 2.2. To simplify the comparisons we focus on the GPV-HMRF method, which Cuadra et al. (2005) found to be the best or nearly the best at noise levels of 5% and higher.

To compare mis-classification rates we use another form of rounding for the results of subvoxel method to produce one of five classes for each voxel. First we compute p_{ij} , the posterior mean proportion of each tissue type j within each whole voxel i from the subvoxel results. Second let $c_i = \sum_j (2j - 1) * p_{ij}$ and round c_i to an integer as the final classification result.

From Figure 10 the five-class rounded subvoxel method is superior to GPV-HMRF at all RF and noise levels considered.

We also compared the average mean square errors between the estimated tissue distributions in each voxel produced by the subvoxel and GPV-HMRF methods to the “fuzzy” tissue membership information for the BrainWeb data. For the GPV-HMRF approach voxels assigned to GW were considered to estimate the voxel’s distribution as 50% GM and 50% WM, with the CG category treated analogously. The results are shown in Figure 11.

Again the more detailed estimates provided by the higher resolution subvoxel approach result in better estimates across all RF and noise levels considered.

5 DISCUSSIONS AND FUTURE WORK

In the previous section, for the prior distributions of \mathbf{z} we used a six-neighbor structure and the simple Potts model for the pure voxel method and the repulsion Potts model with weights equal to 1 for the subvoxel method. We explored some other possible priors on the rf20pn7 data and obtained similar or worse results. A more thorough study of prior models under different neighborhood structures and weightings might be useful to find the best choice of priors for our method.

In classification, one question is how to incorporate more information into the building of classification rules. Most methods proposed for MRI tissue classification just use the T1 weighted image of a subject. But sometimes we can get not only T1 but T2 and PD weighted images as well. To take T2 and PD images into account in addition to T1 might be beneficial.

The inverse temperature parameters β and β_1 and β_2 in the repulsion Potts model are fixed by the pilot study on the BrainWeb data. This works well for data of comparable resolution, noise levels, and intensity non-uniformity to the BrainWeb data. Joint estimation

of these parameters may be necessary to accommodate a wider range of different data sets.

Gibbs sampling is the method we adopt in updating the component indices. Combining the Gibbs sampler with other sampling algorithms and implementations could be helpful to improve mixing of the Markov chain.

Several extensions of the proposed method are worth exploring. It may be possible to extend the approach to support identifying pathologies, such as tumors and lesions. Another possible extension is to spatio-temporal modeling of fMRI data.

References

- Alston, C. L., Mengersen, K. L., Thompson, J. M., Littlefield, P. J., Perry, D., and Ball, A. J. (2005), “Extending the Bayesian mixture model to incorporate Spatial information in analyzing sheep CAT scan images,” *Australian Journal of Agricultural Research*, 56, 373–388.
- Archibald, R., Chen, K., Gelb, A., and Renaut, R. (2003), “Improving tissue segmentation of human brain MRI through preprocessing by the Gegenbauer reconstruction method,” *NeuroImage*, 20, 489–502.
- Besag, J. (1974), “Spatial Interaction and the Statistical Analysis of Lattice Systems (with discussion),” *Journal of the Royal Statistical Society. Series B (Methodological)*, 36, 192–236.
- Besag, J. (1986), “On the Statistical Analysis of Dirty Pictures (with discussion),” *Journal of the Royal Statistical Society. Series B (Methodological)*, 48, 259–302.
- Bourgine, P. and Nadal, J. P. (eds.) (2004), *Cognitive Economics: An interdisciplinary approach*, Springer.

- Celeux, G., Hurn, M., and Robert, C. P. (2000), “Computational and Inferential Difficulties with Mixture Posterior Distributions,” *Journal of the American Statistical Association*, 95, 957–970.
- Chandra, R., Menon, R., Dagum, L., , and Kohr, D. (2000), *Parallel Programming in OpenMP*, Morgan Kaufmann, San Francisco.
- Cocosco, C. A., Zijdenbos, A. P., and Evans, A. C. (2003), “A fully automatic and robust brain MRI tissue classification method,” *Medical Image Analysis*, 7, 513–527.
- Collins, D. L., Zijdenbos, A. P., Kollokian, V., Sled, J. G., Kabani, N. J., Holmes, C. J., and Evans, A. C. (1998), “Design and construction of a realistic digital brain phantom,” *IEEE Transactions on Medical Imaging*, 17, 463–468.
- Cuadra, M. B., Cammoun, L., Butz, T., Cuisenaire, O., and Thiran, J.-P. (2005), “Comparison and Validation of Tissue modelization and statistical classification methods in T1-weighted MR brain images,” *IEEE Transactions on Medical Imaging*, 24, 1548–1565.
- Feng, D. (2008), “Bayesian Hidden Markov Normal Mixture Models with Application to MRI Tissue Classification,” Ph.D. thesis, The University of Iowa.
- Gelman, A. and Meng, X.-L. (1998), “Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling,” *Statistical Science*, 13, 163–185.
- Green, P. J. and Richardson, S. (2002), “Hidden Markov Models and Disease Mapping,” *Journal of the American Statistical Association*, 97, 1055–1070.
- Harris, G., Andreasen, N. C., Cizadlo, T., Bailey, J. M., Bockholt, H. J., Magnotta, V. A., and Arndt, S. (1999), “Improving tissue classification in MRI: a three-dimensional multi-spectral discriminant analysis method with automated training class selection,” *Journal of Computer Assisted Tomography*, 23, 144–154.

- Hauert, C. and Szabó, G. (2005), “Game theory and physics,” *American Journal of Physics*, 73, 405–414.
- Higdon, D. M. (1998), “Auxiliary variable methods for Markov Chain Monte Carlo with applications,” *Journal of the American Statistical Association*, 93, 585–595.
- Hurn, M. (1997), “Difficulties in the use of auxiliary variables in Markov chain Monte Carlo methods,” *Statistics and Computing*, 7, 35–44.
- Ising, E. (1925), “Beitrag zur Theorie des Ferromagnetismus,” *Zeitschrift für Physik*, 31, 253–258.
- Kamber, M., Shinghal, R., Collins, L., Francis, G. S., and Evans, A. C. (1995), “Model-based 3-D segmentation of multiple sclerosis lesions in magnetic resonance brain images,” *IEEE Transactions on Medical Imaging*, 14, 442–453.
- Kohring, G. A. (1996), “Ising Models of Social Impact: the Role of Cumulative Advantage,” *Journal de Physique I*, 6, 301–308.
- Kwan, R. K.-S., Evans, A. C., and Pike, G. B. (1999), “MRI simulation-based evaluation of image-processing and classification methods,” *IEEE Transactions on Medical Imaging*, 18, 1085–1097.
- Leemput, K. V., Maes, F., Vandermeulen, D., and Suetens, P. (1999), “Automated model-based bias field correction of MR images of the brain,” *IEEE Transactions on Medical Imaging*, 18, 885–896.
- Liang, F. (2006), “Continuous Contour Monte Carlo for Marginal Density Estimation with an Application to Spatial Statistical Model,” Tech. rep., Department of Statistics. Texas A&M University.

- MacKay, D. J. (2003), *Information Theory, Inference and Learning Algorithms*, Cambridge University Press.
- Møller, J., Pettitt, A. N., Reeves, R., and Berthelsen, K. K. (2006), “An efficient Markov chain Monte Carlo method for distributions with intractable normalising constants,” *Biometrika*, 93, 451–458.
- Potts, R. B. (1953), “Some generalized order-disorder transformations,” in *Cambridge Philosophic Society*, vol. 48, pp. 106–109.
- Richardson, S. and Green, P. J. (1997), “On Bayesian Analysis of Mixtures with an Unknown Number of Components (with discussion),” *Journal of the Royal Statistical Society Series B-Statistical Methodology*, 59, 731–792.
- Santago, P. and Gage, H. D. (1993), “Quantification of MR Brain Images by Mixture Density and Partial Volume Modeling,” *IEEE Transactions on Medical Imaging*, 12, 566–574.
- Santago, P. and Gage, H. D. (1995), “Statistical models of partial volume effect,” *IEEE Transactions on Medical Imaging*, 4, 1531–1540.
- Schroeter, P., Vesin, J.-M., Langenberger, T., and Meuli, R. (1998), “Robust parameter estimation of intensity distributions for brain magnetic resonance images,” *IEEE Transactions on Medical Imaging*, 17, 172–186.
- Shattuck, D. W., Sandor-Leahy, S. R., Schaper, K. A., Rottenberg, D. A., and Leahy, R. M. (2001), “Magnetic Resonance Image Tissue Classification Using a Partial Volume Model,” *NeuroImage*, 13, 856–876.
- Smith, D. and Smith, M. (2006), “Estimation of binary Markov random fields using Markov chain Monte Carlo,” *Journal of Computational and Graphical Statistics*, 15, 207–227.

- Swendsen, R. H. and Wang, J.-S. (1987), “Nonuniversal Critical Dynamics in Monte Carlo Simulations,” *Physical Review Letters*, 58, 86–88.
- Warfield, S. K., Kaus, M., Jolesz, F. A., and Kikinis, R. (2000), “Adaptive, template moderated, spatially varying statistical classification,” *Medical Image Analysis*, 4, 43–55.
- Wilkinson, D. J. (2005), “Parallel Bayesian Computation,” in *Handbook of Parallel Computing and Statistics*, ed. E. J. Kontoghiorghes, pp. 481–512, Marcel Dekker/CRC Press.
- Winkler, G. (2003), *Image Analysis, Random Fields and Dynamic Monte Carlo Methods: A Mathematical Introduction*, Springer-Verlag, second edn.
- Wolff, U. (1989), “Collective Monte Carlo Updating for Spin Systems,” *Physical Review Letters*, 62, 361–364.
- Woolrich, M. W., Behrens, T. E. J., Beckmann, C. F., and Smith, S. M. (2005), “Mixture models with adaptive spatial regularization for segmentation with an application to FMRI data,” *IEEE Transactions on Medical Imaging*, 24, 1–11.
- Zhang, Y., Brady, M., and Smith, S. (2001), “Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm,” *IEEE Transactions on Medical Imaging*, 20, 45–57.

Table 1: Optimal β values for different methods.

Method	β value
ICM	$\beta = 0.4$
HMRF-EM	$\beta = 0.5$
MCMC-voxel	$\beta = 0.7$
MCMC-subvoxels	$\beta_1 = 0.6, \beta_2 = 0.3$

Figure 1. A 2D coronal slice of a T1 image of a brain. The plot helps to visualize the data used for tissue classification.

Figure 2. The density plot of a T1 image. The density plot of a T1 image often show three modes corresponding to three components, which suggests that a mixture model might be appropriate to model the intensity values.

Figure 3. Illustration of a voxel split into eight subvoxels. To address the issue of partial volume effect, instead of the voxel level, we split each voxel into eight subvoxels and model the intensity values at the subvoxel level, based on the assumption that all subvoxels are pure.

Figure 4. Taking advantage of conditional independence. For a four neighbor structure in 2D, and a six neighbor structure in 3D, given the categories of the black cells, the categories of the white cells are independent and vice versa. As a result, the index vector can be updated in two steps: one for the black cells, one for the white cells. This allows the simulation to be vectorized in high level languages like R or Matlab and also can form the basis for a parallel computing strategy.

Figure 5. Mis-classification rates of the subvoxel method and the other four methods based on the pure voxel assumption for the BrainWeb data. We can see that the simple normal mixture model fitted by the EM algorithm does not perform as well as the other methods, and that its performance deteriorates more rapidly as the noise level increases. The other methods perform similarly, and the larger the noise level, the greater the advantage of the Bayesian methods. The subvoxel strategy is the best when RF=20%, though by a small margin.

Figure 6. Sagittal view of true classification and classification by the subvoxel method for a BrainWeb data set. Using the subvoxel method, we have quite accurate classification results.

Figure 7. Average mean squared errors of the subvoxel method and the other four methods based on the pure voxel assumption for the BrainWeb data. It illustrates the ability of the

subvoxel approach to capture the partial volume effect based on the average mean square errors between the tissue distributions within voxels estimated by the different methods and the true “fuzzy” tissue membership information. Again the simple finite mixture model does not do as well as the other pure voxel approaches. As one might expect, the subvoxel approach does markedly better than the pure voxel methods.

Figure 8. Mis-classification rates of the subvoxel method and the other four methods based on the pure voxel assumption for the Iowa data. Compared with other methods, the subvoxel approach is uniformly best across all four subjects from the Iowa Mental Health Clinical Research Center.

Figure 9. The original MR image, manual classification results, and results from the subvoxel method for one slice of the T1 image from the Iowa data. The higher resolution images constructed by the subvoxel strategy are smoother and capture more detail.

Figure 10. Five components mis-classification rates of the subvoxel method and the GPV-HMRF method. The five-class rounded subvoxel method is superior to GPV-HMRF at all RF and noise levels considered.

Figure 11. Averaged mean squared errors of the subvoxel method and the GPV-HMRF method. The more detailed estimates provided by the higher resolution subvoxel approach result in better estimates across all RF and noise levels considered.

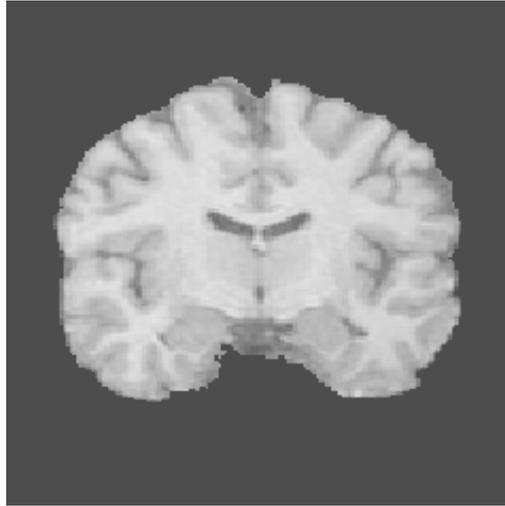


Figure 1: A 2D coronal slice of a T1 image of a brain.

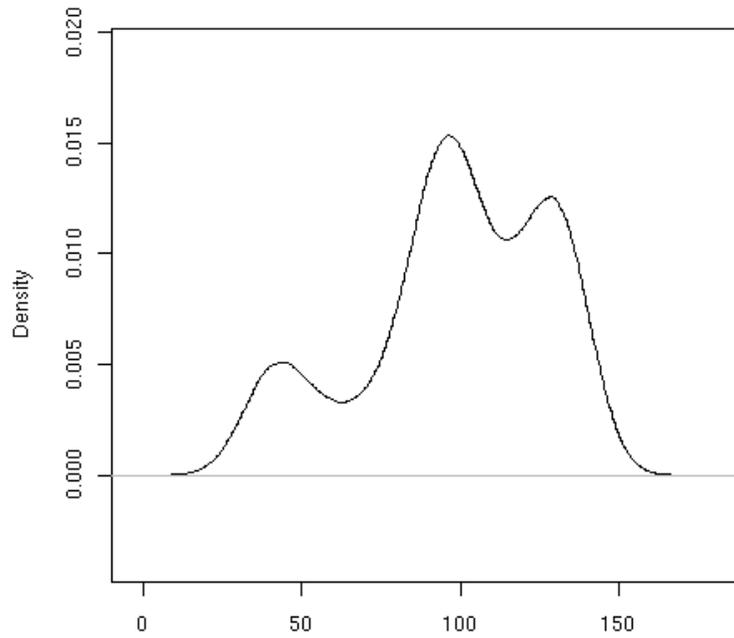


Figure 2: The density plot of a T1 image.

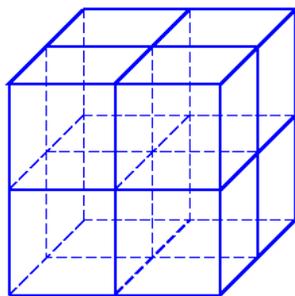
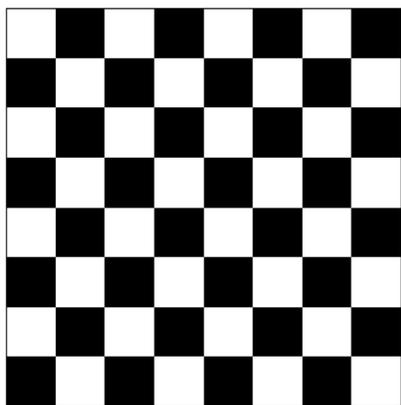
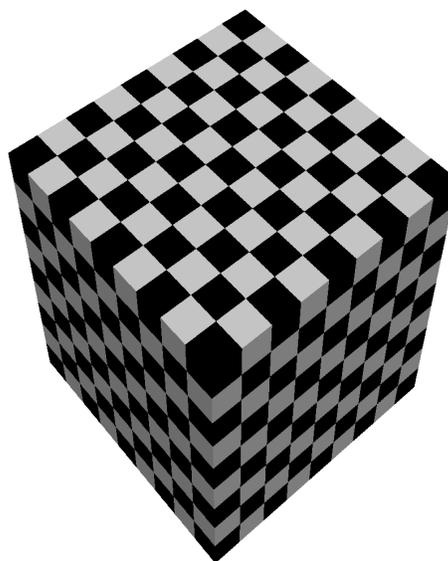


Figure 3: Illustration of a voxel split into eight subvoxels.



(a) Four neighbor configuration in 2D.



(b) Six neighbor configuration in 3D.

Figure 4: Taking advantage of conditional independence.

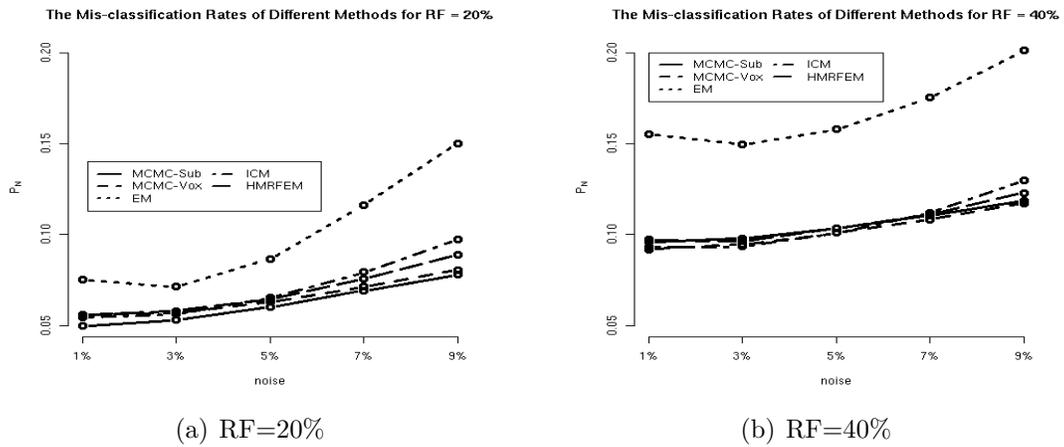


Figure 5: Mis-classification rates of the subvoxel method and the other four methods based on the pure voxel assumption for the BrainWeb data.

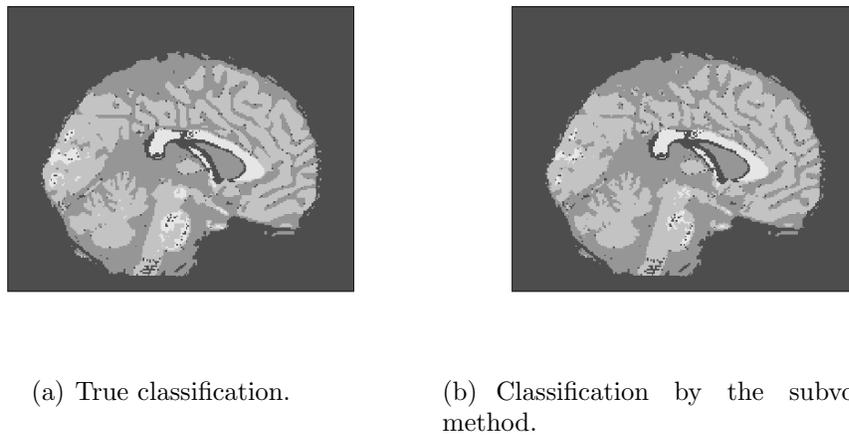


Figure 6: Sagittal view of true classification and classification by the subvoxel method for a BrainWeb data set.

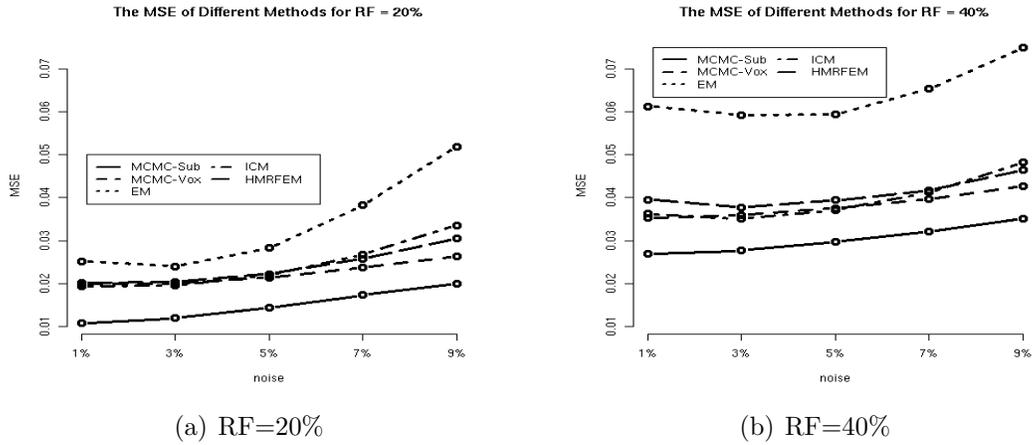


Figure 7: Average mean squared errors of the subvoxel method and the other four methods based on the pure voxel assumption for the BrainWeb data.

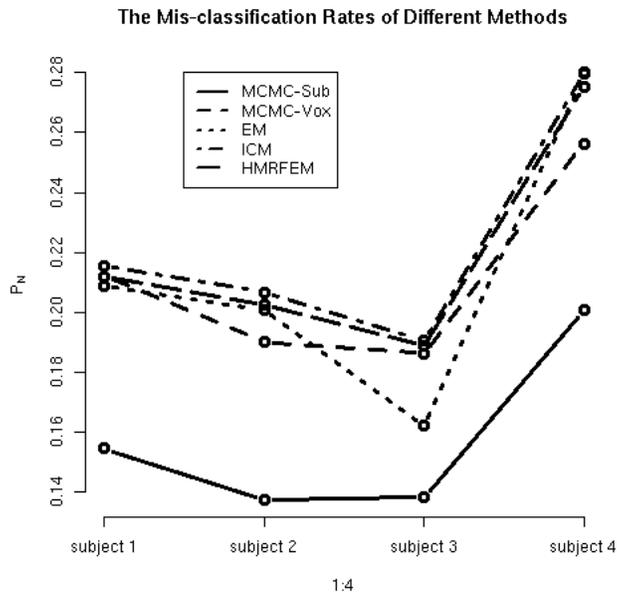
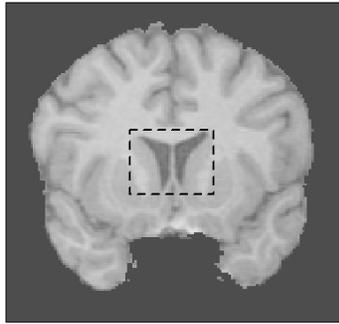
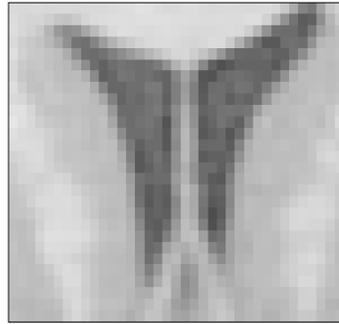


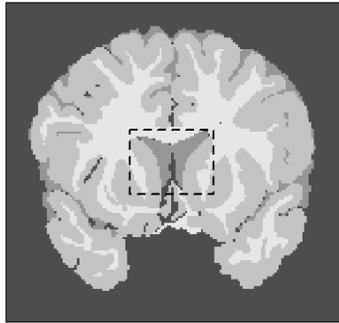
Figure 8: Mis-classification rates of the subvoxel method and the other four methods based on the pure voxel assumption for the Iowa data.



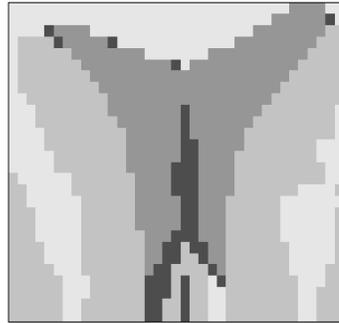
(a) True MR image.



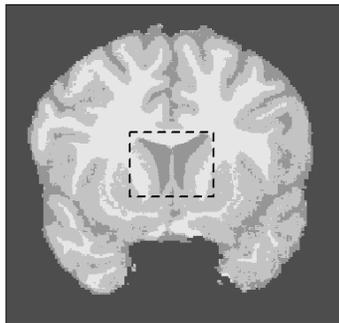
(b) Zoomed true MR image.



(c) Manual classification.



(d) Zoomed manual classification.



(e) Classification by subvoxel method.



(f) Zoomed classification by subvoxel method.

Figure 9: The original MR image, manual classification results, and results from the subvoxel method for one slice of the T1 image from the Iowa data.

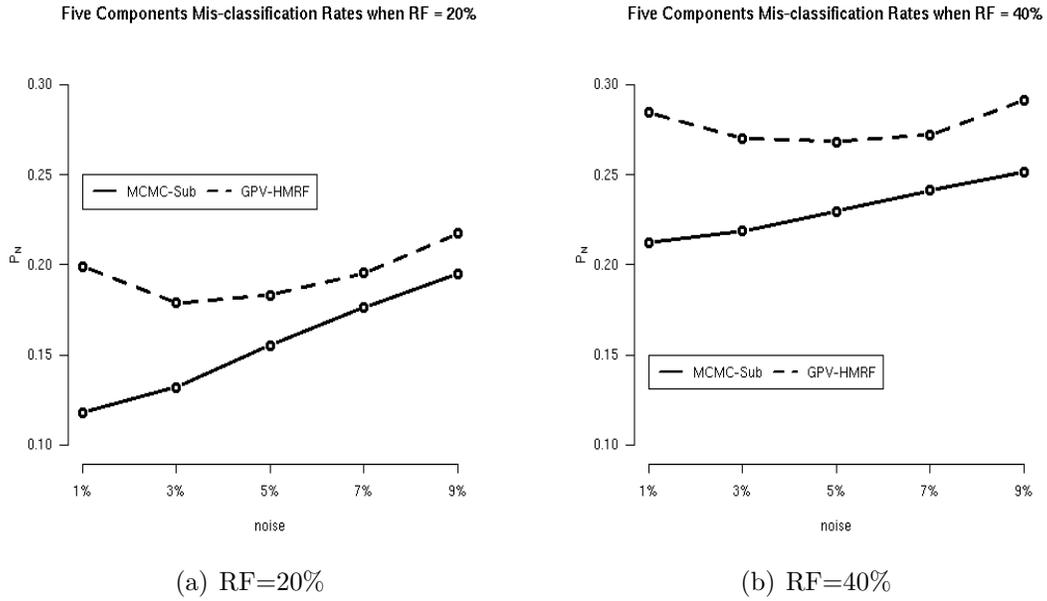


Figure 10: Five components mis-classification rates of the subvoxel method and the GPV-HMRF method.

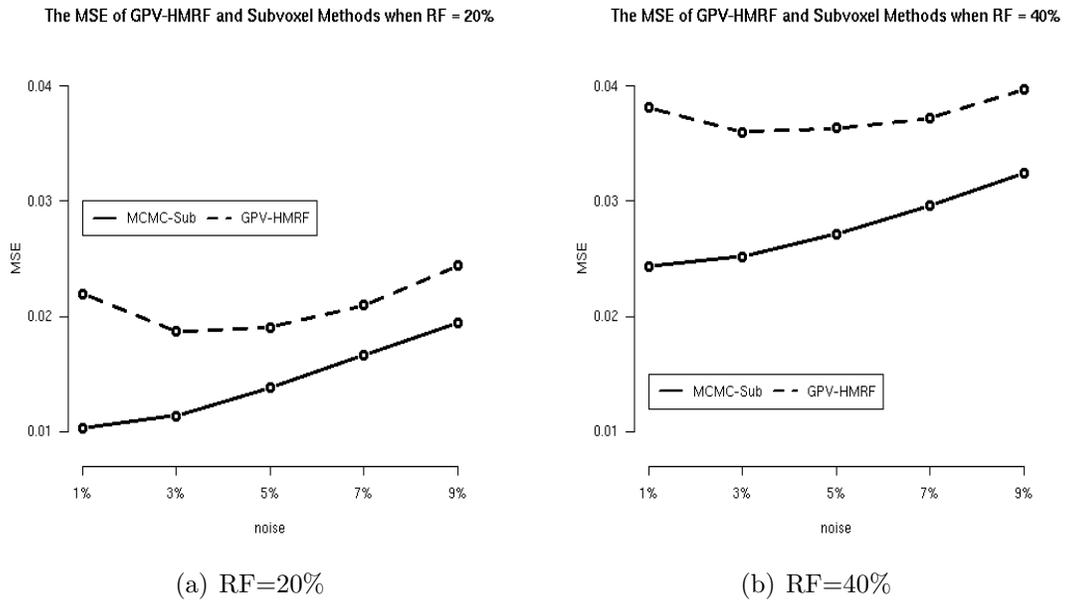


Figure 11: Averaged mean squared errors of the subvoxel method and the GPV-HMRF method.