Bayesian inference and Markov chain Monte Carlo in imaging

D. M. Higdon[a] and J. E. Bowsher[b]

[a]Institute of Statistics and Decision Sciences, Duke University, Durham, USA

[b]Department of Radiology, Duke University, Durham, USA

ABSTRACT

Over the past 20 years, many problems in Bayesian inference that were previously intractable can now be fairly routinely dealt with using a computationally intensive technique for exploring the posterior distribution called Markov chain Monte Carlo (MCMC). Primarily because of insufficient computing capabilities, most MCMC applications have been limited to rather standard statistical models. However, with the computing power of modern workstations, a fully Bayesian approach, with MCMC, is now possible for many imaging applications. Such an approach can be quite useful because it leads not only to "point" estimates of an underlying image or emission source, but it also gives a means for quantifying uncertainties regarding the image. This paper gives an overview of Bayesian image analysis and focuses on applications relevant to medical imaging. Particular focus is on prior image models and outlining MCMC methods for these models.

Keywords: Bayesian inference, Markov chain Monte Carlo, posterior simulation

1. INTRODUCTION

Bayesian image analysis finds its beginnings in papers by Grenander[1] and Geman and Geman.[2] In a fairly typical problem, there exists some true, but unknown scene x, particular features of which must be estimated from observed data y. The data, which may come from more than a single source, provide only partial information about x. In the Bayesian paradigm, one's prior beliefs about plausible scenes x are represented with a probability distribution $\pi(x|\theta)$, which typically depends on additional hyperparameters θ. The data y is then used to update the prior through the likelihood function $L(y|x)$, producing a posterior density

$$\pi(x|y,\theta) \propto L(y|x)\pi(x|\theta)$$

from which inferences about the true scene x are based. Often, the prior $\pi(x|\theta)$ is constructed to ensure the presence of local regularities that are expected in the true scene, as with simple Markov random fields. Such priors primarily serve to massage the likelihood into a more acceptable form. More recently, priors have been developed which more explicitly capture features of the known scene (Refs. 3, 4, 5, 6); such priors often involve anatomical templates or landmarks.

Until recently, the typical end result of a Bayesian formulation is a point estimate of the true scene x, usually taken to be the mode of the posterior distribution

$$\hat{x} = \arg\max_x L(y|x)\pi(x|\hat{\theta}),$$

with $\hat{x}$ depending on estimates $\hat{\theta}$ of the hyperparameters. However, the resulting posterior distribution can give point estimates $\hat{x}$ – which are not conditional on an estimated value for the hyperparameters θ – as well as give uncertainty estimates regarding the true scene x, while accounting for uncertainty in the hyperparameter vector θ. In this fully Bayesian approach, a prior distribution $\pi(\theta)$ must be specified for θ as well. In such applications, the posterior density is usually sufficiently complicated so that it cannot be dealt with analytically. However, samples can often be drawn from the posterior density using Markov chain Monte Carlo (MCMC), giving a straightforward method for exploring the joint posterior distribution $\pi(x,\theta|y)$.

The following section outlines methodology for image analysis under the Bayesian paradigm and also shows how MCMC can be used to draw samples from the resulting posterior distribution. A very simple example is first considered to demonstrate basic methodology. Next, applications to more realistic imaging problems demonstrate the usefulness of this methodology, as well as some limitations. The paper ends with a brief discussion of the potential of such methodology in medical imaging.

Further author information:
E-mail: higdon@stat.duke.edu

2. BASIC METHODOLOGY

This section describes and demonstrates the Bayesian approach to image analysis and shows how MCMC can be used to sample from the resulting posterior distribution. A rather general recipe for constructing the posterior distribution is as follows:

- Specify the likelihood model for the records y given the underlying scene x: $L(y|x)$.
- Specify a prior distribution for the underlying scene x. Typically this will involve additional parameters θ, often called hyperparameters: $\pi(x|\theta)$.
- Specify a prior density for θ: $\pi(\theta)$.
- Form the posterior distribution, which is proportional to the product of the likelihood and the priors

$$\pi(x, \theta|y) \propto L(y|x)\pi(x|\theta)\pi(\theta).$$

- Obtain draws $(x, \theta)^1, \ldots, (x, \theta)^T$ from $\pi(x, \theta|y)$ via MCMC and base posterior inference on this sample.

The above recipe is demonstrated in Section 2.2. But first a few brief details on MCMC are given.

2.1. Markov Chain Monte Carlo

Since analytical determination of posterior quantities such as posterior probabilities and credible intervals are difficult to obtain through standard numerical routines, we use MCMC to estimate these quantities numerically. Detailed descriptions of the underlying methodology for MCMC can be found in Refs.: 7, 8,9, 10,11, and 12, including the accompanying discussions.

Generally let $\pi(z)$ denote the multivariate density of the random variable $Z = (Z_1, \ldots, Z_m)^T$, with $\pi(z_k)$ the corresponding marginal density of Z_k and $\pi(z_k|z_{-k})$ the conditional density for Z_k given $Z_{-k} = (Z_1, \ldots, Z_{k-1}, Z_{k+1}, \ldots, Z_m)$. In practice, each Z_k is often, but not always, univariate. MCMC gives a means for constructing a partial realization $z^1, z^2, \cdots, z^M$ from a Markov chain with limit distribution $\pi(z)$. From this, functionals of π, $T(\pi)$, are estimated by $T(\hat{\pi}_M)$, the functional applied to the resulting empirical distribution. The partial realization can be constructed by stepping through each site k in turn, and updating the value z_k, while leaving the remaining components z_{-k} unchanged.

Updating z_k depends on its *full conditional distribution* $\pi(z_k|z_{-k})$. There are a number of ways to update each component, with Gibbs[13] or Hastings[14] steps being the most commonly used. Updating the kth component with a Gibbs step simply consists of replacing z_k by z'_k drawn from its full conditional distribution. Clearly, if z is a draw from the stationary distribution π, then so to is z', which is identical to z except that its kth component is now z'_k. Hastings steps are more general and do not require draws from the full conditional. In updating component k, a proposal value z^*_k is drawn from an arbitrary density $q_k(\cdot|z)$ which may depend on the current state of the chain z. The ratio of the full conditional posterior densities $\pi(z^*_k|z_{-k})/\pi(z_k|z_{-k})$ is computed, as well as the ratio $q_k(z_k|z^*)/q_k(z^*_k|z)$, where z^* is identical to z except that the kth component is replaced by z^*_k. The updated value for z_k becomes z^*_k with probability

$$r = \min\left\{1, \frac{\pi(z^*_k|z_{-k})q(z_k|z^*)}{\pi(z_k|z_{-k})q(z^*_k|z)}, \right\},$$

otherwise z_k takes on its previous value. The special case where $q(z_k|z^*) = q(z^*_k|z)$ corresponds to a Metropolis[15] update; here the acceptance probability r does not depend on $q(\cdot|\cdot)$. For example, proposals for which $z^*_k \sim N(z_k, \sigma^2)$ or $z^*_k \sim U[z_k - r, z_k + r]$, where σ and r are constants, satisfy this property. There is no need to use the same updating method for each site. In fact, the simple MRF example below uses both Gibbs and Metropolis updates.

Because one usually cannot start with a value z^0 from the prescribed distribution $\pi(z)$, an initial point is chosen and the Markov chain is generated from that point. The chain is allowed to "burn-in" until the stationary distribution is reached. Also, due to limitations of computing space, one may only wish to save every 10^{th} or 100^{th} realization for constructing the empirical distribution $\hat{\pi}$. The Monte Carlo standard errors of estimates $T(\hat{\pi})$ may be estimated using standard time series methods or with other techniques (See Refs. 14, 10, 8, 16).

2.2. Simple application

As a simple example, we consider a computer simulation of a one-dimensional object which is discretized into an array of $n = 40$ sites or pixels indexed by $i = 1, \ldots, n$. The emission intensity for sites $i = 10, \ldots, 30$ is elevated relative to the background emission rates of the remaining sites. Over the observation period, the number of emissions from each site i follows a Poisson distribution with mean x_i where $x_i = 20$ counts for the "hot" interior sites, and $x_i = 10$ counts for the background exterior sites as shown by the solid line in Figure 1.

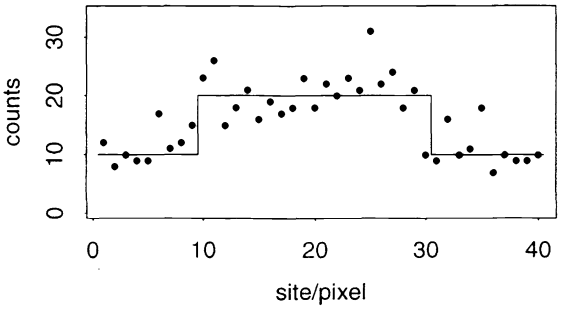

Figure 1. Simulated counts from a one-dimensional emission source; the true emission intensity is given by the solid line.

Hence the likelihood for the data y given the underlying count intensities x is given by

$$L(y|x) \propto \prod_{i=1}^{n} e^{-x_i} x_i^{y_i}, \quad \text{all } x_i \geq 0.$$

where

$$x_i = \begin{cases} 20 & \text{if } 10 \leq i \leq 30 \\ 10 & \text{if } i < 10 \text{ or } i > 30 \end{cases}.$$

Since the purpose of this data is to demonstrate basic methodology, more realistic features such as blurring are not considered here. Under the Bayesian paradigm, specification of the prior distribution of the object intensities x is required. Two different formulations are given below: one using a simple MRF prior; and the other which uses a template for the emission object.

MRF based prior

A MRF based prior serves to inject regularity in object intensity x by requiring nearby pixel intensities to be similar. Perhaps the simplest such model is a Gaussian random walk with

$$x_i | x_{i-1} \sim N(x_{i-1}, 1/\theta)$$

where θ controls the scale of the random walk increments. In this case, the full conditional for any interior x_i has the form

$$x_i | x_{-i} \; N\left(\frac{x_{i+1} + x_{i-1}}{2}, \frac{1}{2\theta}\right);$$

for the two edge sites, the full conditional is centered at its one neighboring value, with variance $1/\theta$. Multivariately, the prior for x has the form

Template based prior

We construct a template prior to correspond to our belief that the emitting source consists of a "hot" region and a background region, each with it's own constant intensity. This template may be parameterized by the 4 variables:

- x_h - the emission intensity of the hot region;
- x_b - the emission intensity of the background region;
- i_ℓ - leftmost site of the hot region;
- i_r - rightmost site of the hot region.

Hence the emission intensity of the object is a deterministic function of these 4 parameters, with the intensity at site i given by

$$x_i = \begin{cases} x_h & \text{if } i_\ell \leq i \leq i_r \\ x_b & \text{if } i < i_\ell \text{ or } i_r \geq i \end{cases}. \tag{1}$$

Specifying a uniform prior for (x_h, x_b, i_ℓ, i_r) subject to the above conditions gives the joint prior density

$$\pi(x|\theta) \propto \theta^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\theta \sum_{i=1}^{n-1}(x_i - x_{i+1})^2\}.$$

A gamma prior is specified for the hyperparameter θ so that

$$\pi(\theta) \propto \theta^{a-1} \exp\{-b\theta\},\ \theta > 0.$$

This leads to the posterior distribution

$$\pi(x, \theta|y) \propto \prod_{i=1}^{n} e^{-x_i} x_i^{y_i} \times \theta^{\frac{n}{2}+a-1} \times \exp\{-\tfrac{1}{2}\theta \sum_{i=1}^{n-1}(x_i - x_{i+1})^2 - b\theta\},$$
$$\text{all } x_i \geq 0,\ \theta > 0.$$

The resulting full conditionals are then

$$\pi(x_i|\cdots) \propto e^{-x_i} x_i^{y_i} \times \exp\{-\tfrac{1}{2}\theta[(x_i - x_{i-1})^2 + (x_i - x_{i+1})^2]\},\ x_i \geq 0$$
$$\pi(\theta|\cdots) \propto \theta^{\frac{n}{2}+a-1} \exp\{-\tfrac{1}{2}\sum_{j=1}^{n-1}(x_j - x_{j+1})^2 - b\theta\}, \theta > 0$$

After initializing x at $\frac{1}{2}(y + \bar{y})$, and θ with a draw from its full conditional, each parameter was updated in turn according to its full conditional distribution. Metropolis updates were used for each x_i; the candidate was drawn from a uniform distribution over $(x_i \pm 3)$. The hyperparameter θ was updated with a Gibbs step (ie. a draw from its full conditional). Several realizations of x from the MCMC output are shown in Figure 2a. The MCMC chain was run for 1,000 burn-in cycles, and then realizations were recorded for the next 10,000 cycles. The collected realizations can be used, for example, to obtain pointwise 90% credible regions for the true image x as shown below in Figure 3a.

$$\pi(x_h, x_b, i_\ell, i_r) \propto 1_{[1 \leq i_\ell < i_r \leq n; x_b > 0; x_h > 0]}.$$

This leads to the posterior distribution

$$\pi(x_h, x_b, i_\ell, i_r|y) \propto \prod_{i=1}^{n} e^{-x_i} x_i^{y_i} \times 1_{[1 \leq i_\ell < i_r \leq n; x_b > 0; x_h > 0]}$$

where the intensities x_i are determined by (1).

The resulting full conditionals are then

$$\pi(x_h|\cdots) \propto \prod_{i=1}^{n} e^{-x_i} x_i^{y_i} \cdot 1_{[x_h > 0]}$$
$$\pi(x_b|\cdots) \propto \prod_{i=1}^{n} e^{-x_i} x_i^{y_i} \cdot 1_{[x_b > 0]}$$
$$\pi(i_\ell|\cdots) \propto \prod_{i=1}^{n} e^{-x_i} x_i^{y_i} \cdot 1_{[0 \leq i_\ell < i_r]}$$
$$\pi(i_r|\cdots) \propto \prod_{i=1}^{n} e^{-x_i} x_i^{y_i} \cdot 1_{[i_\ell \leq i_r < n]}$$

After initializing (x_h, x_b, i_ℓ, i_r) at $(20, 10, 10, 30)$, each parameter was updated in turn according to its full conditional distribution. Metropolis updates were used for each of the parameters: the candidate value for x_h was drawn uniformly over the range $(x_h \pm 2)$; the candidate value for x_b was drawn uniformly over the range $(x_b \pm 2)$; the candidate value for i_ℓ was drawn uniformly over the six points $\{i_\ell \pm 3, i_\ell \pm 2, i_\ell \pm 1\}$; and the candidate value for i_r was drawn uniformly over the six points $\{i_r \pm 3, i_r \pm 2, i_r \pm 1\}$. The MCMC chain was run for 1,000 burn-in cycles, and then realizations were recorded for the next 10,000 cycles. Several realizations of x from the MCMC output are shown below in Figure 2b. The collected realizations can be used, for example, to obtain pointwise 90% credible regions for the true image x as shown below in Figure 3b.

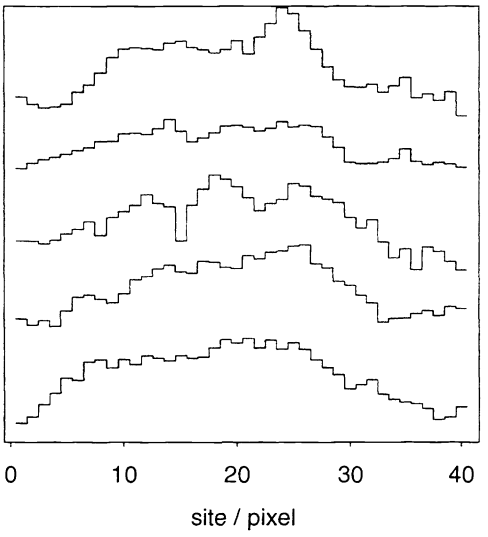

Figure 2a. Posterior realizations of the emission intensities.

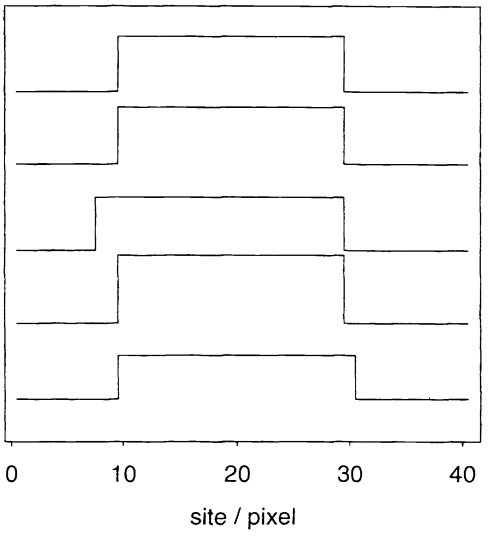

Figure 2b. Posterior realizations of the emission intensities.

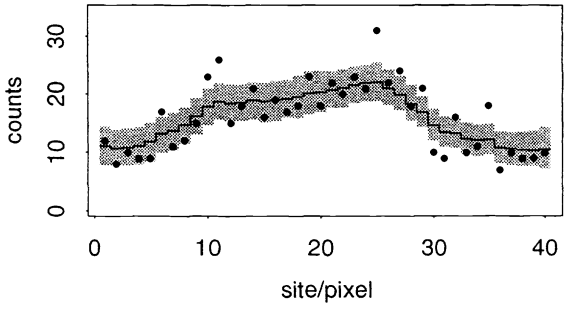

Figure 3a. Posterior mean estimate (solid line) and 90% credible region for the emission intensities (shaded area).

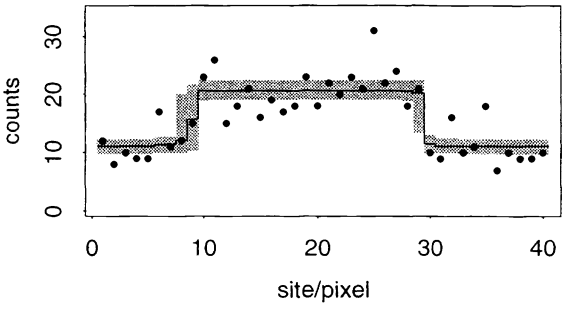

Figure 3b. Posterior mean estimate (solid line) and 90% credible region for the emission intensities (shaded area).

Because the two formulations above use rather different priors for x, the nature of the resulting posterior distributions for x differs a fair bit. This simple example also shows that even rather uninformative data can lead to a sharp posterior for x if sufficient information is built into the prior, as with the template formulation.

3. APPLICATIONS

3.1. 2-d SPECT and PET Reconstructions

This section considers two examples of 2-dimensional ECT reconstructions: one SPECT; and the other PET. A standard Poisson model for the ECT data is assumed. Following Vardi, Shepp and Kaufman,[17] let p_i^t denote the probability that a positron or photon emitted from pixel i results in a registration at detector or tube t. In the case of PET, let x_i denote the mean intensity of emission of positrons from pixel i over the course of the study; in the case of single photon emission computed tomography (SPECT) let this variable denote the mean emission rate of photons from pixel i. Let y_t denote the observed number of registrations in tube or detector t. Assuming that the probabilities $\{p_i^t\}$ are known, the likelihood function $L(\cdot)$ may be expressed

$$L(y|x) \propto \prod_{t=1}^{T} \exp\left(-\sum_i p_i^t x_i\right)\left(\sum_i p_i^t x_i\right)^{y_t}. \tag{2}$$

First a reconstruction of a cylindrical physical SPECT phantom is considered. As a prior for the emission intensities of a cross section of the phantom we use a 2-dimensional extension of the Gaussian MRF prior given in

the previous example

$$\pi(x|\theta) \propto \theta^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\theta \sum_{i\sim j}(x_i - x_j)^2\} \tag{3}$$

where the sum is over each vertical and horizontal adjacency in the 2-dimensional lattice of sites in x. A gamma prior is used for the hyperparameter θ: $\pi(\theta) \propto \theta^{a-1}e^{-b\theta}$. The values $a = 1$ and $b = .005$ are used here, though the resulting posterior is rather insensitive to a fairly wide range of sensible choices for a and b.

The above formulation leads to the posterior distribution

$$\pi(x,\theta|y) \propto \prod_{t=1}^{T} \exp\left(-\sum_i p_i^t x_i\right)\left(\sum_i p_i^t x_i\right)^{y_t} \times \theta^{\frac{n}{2}} \exp\{-\tfrac{1}{2}\theta \sum_{i\sim j}(x_i - x_j)^2\} \times \theta^{a-1}e^{-b\theta}$$

and yields the full conditionals

$$\begin{aligned} \pi(x_i|x_{-i},\theta,y) &\propto \prod_{t=1}^{T} \exp\left(-\sum_i p_i^t x_i\right)\left(\sum_i p_i^t x_i\right)^{y_t} \cdot \exp\{-\tfrac{1}{2}n_i\theta(x_i - \bar{x}_{\partial i})^2\} \\ \pi(\theta|x,y) &\propto \theta^{a+\frac{n}{2}-1} \exp\{b + \tfrac{1}{2}\sum_{i\sim j}(x_i - x_j)^2\} \end{aligned}$$

where n_i is 4 – the number of sites horizontally or vertically adjacent to i – and $\bar{x}_{\partial i}$ is the average of these four neighboring sites.

Hastings updates are used for x_i drawing the canditate value x_i^* from a gamma distribution with mean and variance both equal to x_i so that

$$q(x_i^*|x_i) = \frac{x_i^{*x_i - 1}e^{-x_i^*}}{\Gamma(x_i)}.$$

The candidate x_i^* is then accepted with probability

$$\min\left\{1, \frac{\pi(x_i^*|x_{-i},\theta,y)q(x_i|x_i^*)}{\pi(x_i|x_{-i},\theta,y)q(x_i^*|x_i)}\right\}.$$

Though this choice of gamma distribution works well in this application, the candidate distribution may require additional tuning depending on the spread in the full conditional for x_i. An alternative might be to use a gamma distribution with mean x_i and variance βx_i, where β is a constant chosen so that the candidate draw is accepted roughly half the time. The parameter θ may be updated using a Gibbs step, sampling directly from a gamma distribution with shape parameter $a + \frac{n}{2}$ and scale parameter $1/(b + \frac{1}{2}\sum_{i\sim j}(x_i - x_j)^2)$. Additional details on MCMC sampling in PET and SPECT applications may be found in Refs. 18 and 19. For the SPECT application, the posterior mode and mean for x are shown in Figure 4. The rightmost frame shows a 90% credible interval for a cross-section of x.

Since the prior in (3) sometimes yields overly smooth reconstructions, other MRF priors have been proposed that allow large intensity differences between adjacent pixels. One such prior was proposed by Geman and McClure[20] which has the form

$$\pi(x|\theta) \propto \theta_1^{\frac{n}{2}} \exp\left\{-\tfrac{1}{2}\theta \sum_{i\sim j} \frac{1}{1 + [(x_i - x_j)/\sigma]^2}\right\}, \quad 0 \le x_i \le M, \text{ for all } i,$$

where σ and M are specified constants. Since the value of each x_i is constrained to lie between 0 and M, this is a proper prior distribution. However, because the density does not go to 0 as $|x_i - x_j|$ goes to ∞ the prior allows occasional large intensity "jumps" in x. Such priors can yield very satisfactory point estimates obtained from a posterior mode, however the resulting posterior distribution may be so spread out as to give useless information regarding uncertainty. The posterior realizations shown in Figure 5 were obtained by using the prior of Geman and McClure in a PET application. That the posterior realizations of x are so noisy and irregular suggests that the prior information, when combined with the data, is not insufficiently strong to give sensible realizations.

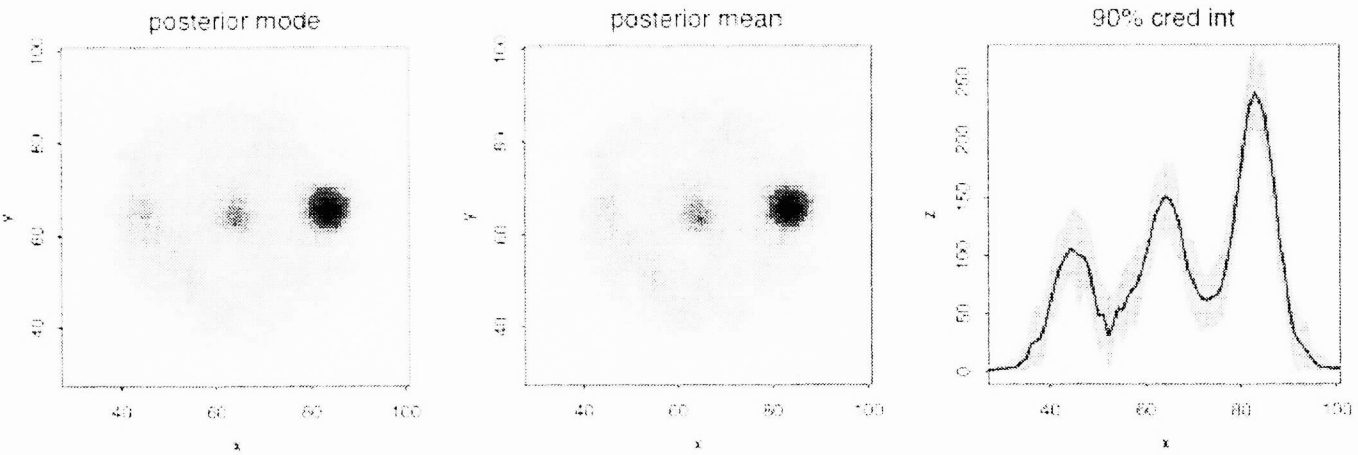

Figure 4. Posterior summary of a SPECT reconstruction of a physical phantom. The rightmost frame gives a pointwise 90% credible interval for a slice along the row of pixels corresponding to y= 64.

In hindsight this isn't surprising since the MRF gives rather crude and local information regarding the true scene x. In the case of relatively low count SPECT and PET imaging, there is insufficient information in the data to overcome the shortcomings in the prior. Hence for such imaging modalities more informative prior information is required before the posterior distribution can be trusted to realistically quantify uncertainty. High-level priors, such as those given in Refs. 21, 4, 3, 22 and 5 may fare much better. Another alternative is to use additional information from other imaging modalities as shown in final example below.

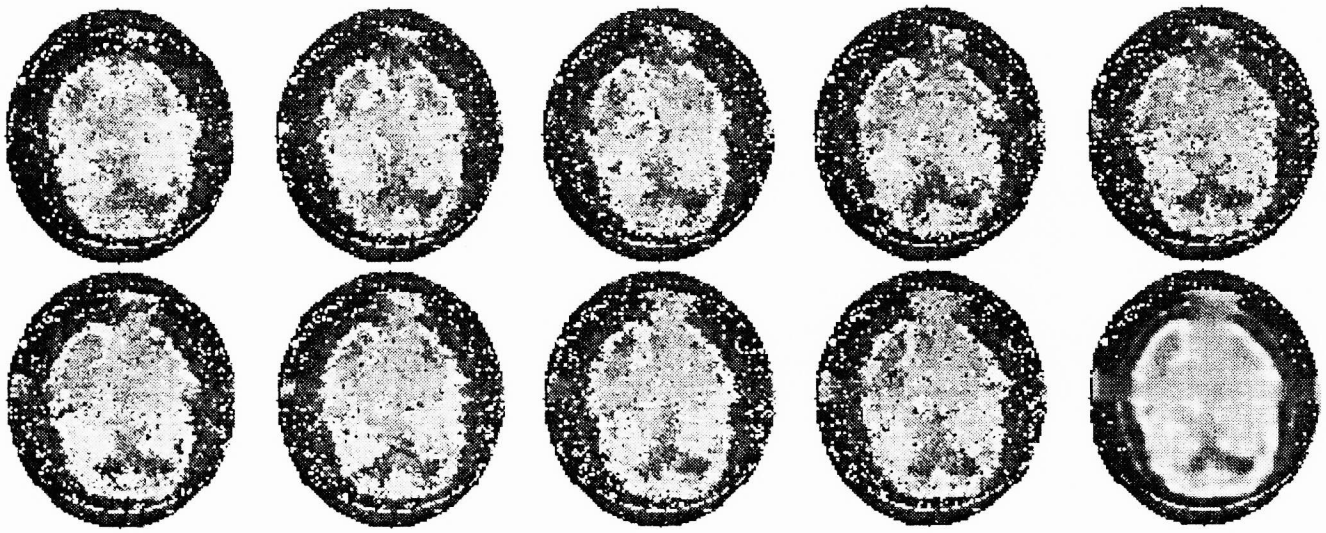

Figure 5. Posterior realizations of a PET reconstruction using a insufficiently restrictive prior. The final frame shows the posterior mean estimate.

3.2. Combining information from two sources

This final example is a computer simulation study which combines information from MRI and PET. The MRI phantom, shown in the left-hand frame of Figure 6, was obtained from the Hoffman brain phantom[23] by adding a dark sphere representing surgically resected tissue and by adding a surrounding shell representing high gadolinium uptake and a breakdown in the blood brain barrier. The PET phantom is the same as the MRI phantom except that the upper left portion of the shell has an activity 50% above that of gray matter, as might occur in PET-FDG imaging with a fairly high grade tumor. The lower right portion of the shell has the same activity as white matter.

Figure 6 Two-dimensional MRI and PET phantoms used for the computer simulation study of Section C.1. The left-hand image represents an MRI. It includes a sphere from which tumor has been resected and a surrounding bright shell of high gadolinium uptake. The right-hand image is the PET phantom. The upper left portion of the MRI-defined shell has an activity 50% above that of gray matter, which is characteristic of high grade tumor on a FDG PET scan. The lower right portion of the shell has the same activity as white matter.

As a prior distribution for photon emission for the PET data, an "anatomical" prior is used which is based on an a priori segmentation of the MRI scan. In the spirit of the "template" prior in Section 2.2, x is a piecewise constant over k connected regions. In this prior, x must be a refinement of the partition obtained from segmenting the MRI scan - so x must have at least as many separate regions as in the MRI segmentation, and possibly more. The number of distinct regions is treated here as a parameter. Thus x must contain regions corresponding to anatomical tissue types and possibly additional regions as determined by the PET data. Of particular importance here is detecting a separate region within the spherical shell-shaped lesion having high gadolinium uptake which cannot be detected with the MRI scan.

Conditional on the number of separate regions k, the prior for x is controlled by an image of region labels r which may take on values $1, \ldots, k$ and the region intensity means $\mu_1, \ldots, \mu_k$. The prior for the labels r is uniform subject to the constraint that the resulting partition must be a refinement of the MRI partition; each of the μ_k's are given independent uniform priors over $[0, M]$ where M is fixed. And finally, the overall number of levels k is given a Poisson prior as is used in non-spatial classification modeling in Ref. 24. This gives a prior of the form $\pi(\mu|k) \times \pi(r|k) \times \pi(k)$; the intensities are then a deterministic function of these parameters with $x_i = \mu_{r_i}$. Because the dimension of the posterior distribution may vary, depending on k, generating posterior realizations is not as straightforward as in the previous examples. Sampling from such variable dimension distributions requires more sophisticated techniques such as the *reversible jump* metropolis algorithm of Green[25,24] - no attempt to explain this methodology is made here.

The resulting posterior favors the formation of a second region within the shell of high gadolinium uptake corresponding to the high grade tumor. Posterior draws for the intensity and the number of pixels corresponding to the high grade tumor region are shown in Figure 7. All posterior realizations obtained after the burn-in cycles contain at least 12 pixels of tumor region. Figure 8 shows the actual tumor region and the pixels assigned to this region with posterior probability of 75%, 50%, 25%, 10% and 2% from left to right.

4. DISCUSSION

This paper has demonstrated the use of Bayesian Inference via MCMC in medical imaging applications. With recent advances in computing, such an approach is possible in quite a range of applications, only a small sample of which were considered here. Such advances will likely prove worthwhile in areas where uncertainty plays a key role.

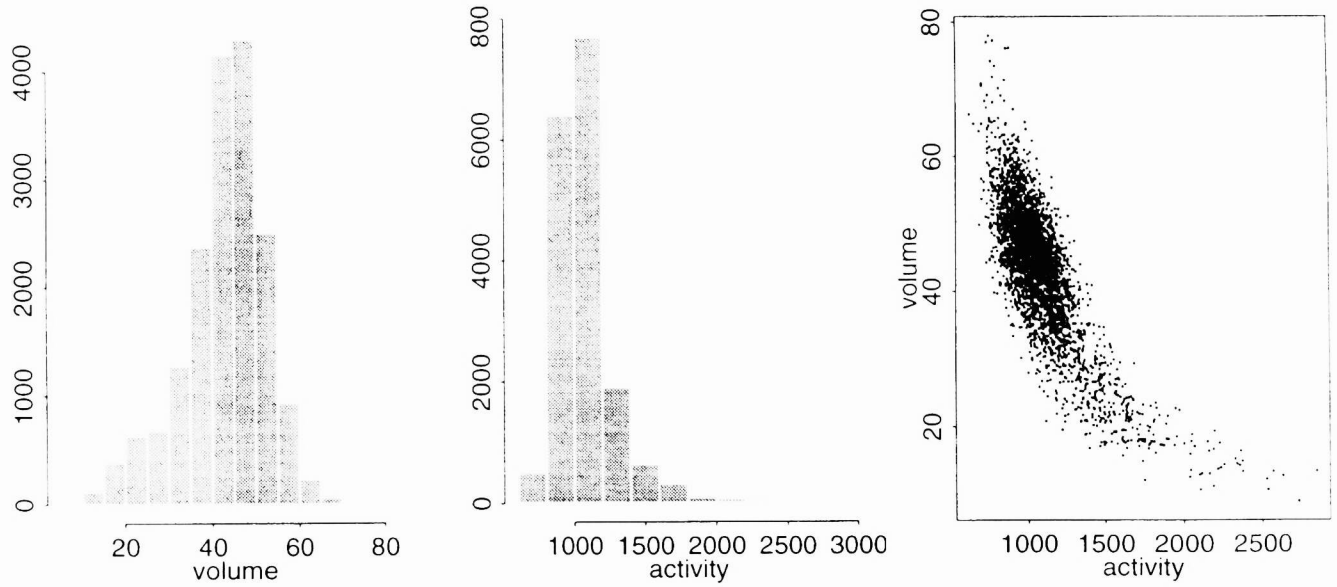

Figure 7. The first two frames are MCMC estimates of posterior distributions for the volume and activity of the tumor shown in Figure 6. The true volume is 0.32 ml, and the true activity is 980 (in thousands of emissions per voxel). The local MAP estimates were 0.13 ml and 1590. The last frame indicates the estimated joint posterior probability distribution for tumor volume and activity, with denser areas indicating higher probability.

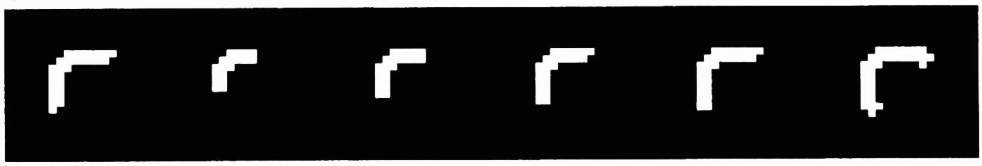

Figure 8. The leftmost frame indicates the voxels that are in the tumor of the PET phantom in Figure 6. Successive frames highlight voxels whose probability of membership in the tumor was estimated to be at least 75%, 50%, 25%, 10%, and 2%, respectively.

Although this is a relatively new approach for imaging related problems, methodology shows promise in classification, feature recognition, and feature location.

Some care must be taken in choosing the prior distribution. Priors which lead to good point estimates (eg. posterior mode solutions) can give unrealistically dispersed posterior distributions as in the second example of Section 3.1. Here any uncertainty estimates derived from such a posterior will not be reliable. This highlights a need to develop realistic priors for such applications.

Acknowledgments

Research was supported by National Science Foundation grant DMS-9704425 and a Whitaker Foundation grant.

REFERENCES

1. U. Grenander, "Tutorial in pattern theory," report, Division of Applied Mathematics, Brown University, 1983.
2. S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, pp. 721–741, 1984.
3. U. Grenander and M. I. Miller, "Representations of knowledge in complex systems (disc: P581-603)," *Journal of the Royal Statistical Society (Series B)* **56**, pp. 549–581, 1994.
4. D. B. Phillips and A. F. M. Smith, "Bayesian faces via hierarchical template modeling," *Journal of the American Statistical Association* **89**, pp. 1151–1163, 1994.

5. J. K. Laading, C. McCulloch, V. E. Johnson, D. R. Gilland, and R. J. Jaszczak, "A hierarchical feature based deformation model applied to 4d cardiac spect data," to appear in *Information Processing in Medical Imaging*, 1999.
6. K. M. Hanson and G. S. Cunningham, "Operation of the Bayes inference engine," to appear in *Maximum Entropy and Bayesian Statistics*, V. Dose et al., ed., 1999.
7. A. F. M. Smith and G. O. Roberts, "Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods," *Journal of the Royal Statistical Society (Series B)* **55**, pp. 3–23, 1993.
8. J. Besag and P. J. Green, "Spatial statistics and Bayesian computation (with discussion)," *Journal of the Royal Statistical Society (Series B)* **16**, pp. 395–407, 1993.
9. W. R. Gilks, D. G. Clayton, D. J. Spiegelhalter, N. G. Best, A. J. McNeil, L. D. Sharples, and A. J. Kirby, "Modelling complexity: Applications of Gibbs sampling in medicine (disc: P53-102)," *Journal of the Royal Statistical Society (Series B)* **55**, pp. 39–52, 1993.
10. C. J. Geyer, "Practical markov chain monte carlo (with discussion)," *Statistical Science* **7**, pp. 473–511, 1992.
11. L. Tierney, "Markov chains for exploring posterior distributions (with discussion)," *Annals of Statistics* **21**, pp. 1701–1762, 1994.
12. J. Besag, P. J. Green, D. M. Higdon, and K. Mengersen, "Bayesian computation and stochastic systems (with discussion)," *Statistical Science* **10**, pp. 3–66, 1995.
13. A. E. Gelfand and A. F. M. Smith, "Sampling based approaches to calculating marginal densities," *Journal of the American Statistical Association* **85**, pp. 389–409, 1990.
14. W. K. Hastings, "Monte Carlo sampling methods using Markov chains and their applications," *Biometrika* **57**, pp. 97–109, 1970.
15. N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller, "Equations of state calculations by fast computing machines," *Journal of Chemical Physics* **21**, pp. 1087–1091, 1953.
16. G. S. Fishman, "An implementation of the batch means method," *INFORMS Journal on Computing* **9**, pp. 196–310, 1997.
17. Y. Vardi, L. Shepp, and L. Kaufman, "A statistical model for positron emission tomography," *Journal of the American Statistical Association* **80**, pp. 8–25, 1985.
18. D. M. Higdon, V. E. Johnson, J. E. Bowsher, T. G. Turkington, D. R. Gilland, and R. J. Jaszczack, "Fully Bayesian estimation of Gibbs hyperparameters for emission computed tomography data," *IEEE Transactions on Medical Imaging* **16**, pp. 516–526, 1997.
19. I. Weir, "Fully Bayesian reconstructions from single photon emission computed tomography," *Journal of the American Statistical Association* **92**, pp. 49–60, 1997.
20. S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bulliten of the International Statistical Institute* **52**, pp. no.4 5–21, 1987.
21. Y. Amit, U. Grenander, and M. Piccioni, "Structural image restoration through deformable templates," *Journal of the American Statistical Association* **86**, pp. 376–387, 1991.
22. J. E. Bowsher, V. E. Johnson, T. G. Turkington, R. J. Jaszczak, C. E. Floyd, and R. E. Coleman, "Bayesian reconstruction and use of anatomical a priori information for emission tomography," *IEEE Transactions on Medical Imaging* **12**, pp. 673–686, 1996.
23. E. Hoffman, P. Cutler, W. Digby, and J. Mazziotta, "3-D phantom to simulate cerebral blood flow and metabolic images for PET," *IEEE Transactions on Nuclear Science* **37**, pp. 616–620, 1990.
24. S. Richardson and P. J. Green, "On Bayesian analysis of mixtures with an unknown number of components (with discussion)," *Journal of the Royal Statistical Society (Series B)* **59**, pp. 731–792, 1997.
25. P. J. Green, "Reversible jump Markov chain Monte Carlo computation and Bayesian model determination," *Biometrika* **82**, pp. 711–732, 1995.
26. D. M. Higdon, "Axiliary variable methods for Markov chain Monte Carlo with applications," *Journal of the American Statistical Association* **93**, pp. 585–595, 1998.