Fully Bayesian reconstructions from single photon emission computed tomography data*

Iain S. Weir
University of Bristol†

Abstract

It is well documented that a Bayesian model with a pairwise difference prior can give far more satisfactory reconstructions of single photon emission computerised tomography data than MLE estimates. However, successful reconstruction requires appropriate prior parameter values. Rather than using point estimates either from training sets or from the data, a fully Bayesian approach is presented where the parameters are considered drawn from hyperpriors. The approach is problematical since the normalisation constant in the prior is an intractable function of its parameters. Here, the constant is estimated off-line by reverse logistic regression. Markov chain Monte Carlo methods are used on simulated, physical phantom, and real data to gain estimates of the posterior mean and pixelwise credibility bounds. The resulting reconstructions are compared to those obtained by filtered back-projection and the one step late solution to the fixed parameter posterior mode.

Keywords: Fully Bayesian; Gibbs prior; Hastings algorithm; Markov chain Monte Carlo; Normalising constant; Reverse logistic regression; Single photon emission computerised tomography.

1 Introduction

Single photon emission computerised tomography (SPECT) is a method of imaging internal physiological events. A radioactively labelled substance that has been

*Work supported by the SERC Complex Stochastic Systems Initiative.

†Correspondence address: I. S. Weir, Department of Mathematics, University of Bristol, Bristol, BS8 1TW, UK. Electronic mail: `Iain.Weir@bristol.ac.uk`

designed to show a physiological function of a particular organ is introduced into the patient, usually through injection or inhalation. The substance is chosen as one known to become concentrated in the organ of interest in a manner related to the phenomenon under study. Photon emission then occurs in the organ at a rate varying spatially according to the concentration. Indirect measurements of the concentration can be obtained by counting emitted photons using a gamma camera that is swung around the patient to collect a sequence of two-dimensional projections at equally spaced angles. The object of *reconstruction* is to extract representations of the pattern of isotope concentration from this sequence of projections.

The data consist of an array of counts of detected photons, which we denote by $\mathbf{y} = \{y_t\}$. Note that t denotes a (projection, pixel) pair. The appropriate statistical model for the data is essentially that which applies to all emission data (see, for example, Shepp and Vardi (1982)). The reconstruction space is subdivided into rectangular pixels denoted by $\mathbf{x} = \{x_s\}$ where x_s is the true total radiopharmaceutical concentration within pixel s and thus non-negative. The observed counts are independent Poisson distributed random variables, with

$$y_t \sim \text{Poisson}\left(\sum_s a_{ts} x_s\right)$$

where a_{ts} corresponds to the mean rate of arrival at detector location t per unit concentration at pixel s. The weights $\{a_{ts}\}$ depend on various factors: the geometry of the detection system, the strength of the isotope and exposure time, and the extent of attenuation and scattering between source and detector. Weir and Green (1994) discuss and compare two methods of constructing these weights: the first utilises only knowledge of the geometric/physical properties of the SPECT system, whilst the second method is an extension of the empirical model of Geman, Manbeck, and McClure (1991). The first method is found to be preferable and shall be used here.

The joint probability function of the data given the isotope concentration is given by

$$p(\mathbf{y}|\mathbf{x}) = \prod_t \frac{\exp(-\sum_s a_{ts} x_s)(\sum_s a_{ts} x_s)^{y_t}}{y_t!} \tag{1.1}$$

and Shepp and Vardi (1982) proposed using the EM algorithm for obtaining its MLE solution. However, MLE solutions have the unattractive feature of being very noisy in appearance because of the high dimensionality of the parameter vector $\mathbf{x}$. Bayesian reconstruction from SPECT data was proposed by Geman and McClure (1985), following the general paradigm of Bayesian image analysis introduced by Geman and Geman (1984), and Besag (1986). This involves the construction of a second model which is a probability distribution on the space of true patterns of isotope concentration. It has become standard practice in research to use a pairwise difference Gibbs prior with a potential function that incorporates a smoothing and scale parameter. Reconstruction can then be accomplished by considering the posterior distribution that follows from Bayes' theorem on the two model components and leads to far more satisfactory results. Various reconstruction methods have been used: Geman and McClure (1985, 1987, 1991 [with Manbeck]) use Iterated Conditional Modes (ICM) and Markov chain Monte Carlo (MCMC) methods, whilst Green (1990) proposed the one step late (OSL) algorithm as an adaptation of the EM algorithm which is impractical for posterior maximisation and uses MCMC (1995).

Regardless of the method of reconstruction, success in obtaining realistic reconstructions relies upon using appropriate prior parameter values. A simple approach is to use training data sets to gain an insight into choosing their values. Weir (1993) uses simulated training sets from a suitably designed phantom and finds the transition from simulated to real data effortless. There is a case for using fixed parameters in clinical practice, to ensure comparability between patients. Also one's prior knowledge could well include specified parameters. However, it is far more appealing and flexible to adopt the empirical Bayesian approach and estimate the parameter values from the data. Geman and McClure (1987) assume that the scale parameter is known, and develop a Gibbs/EM approach to estimate the smoothing parameter from the data. However, this procedure is not straightforward and can be computationally demanding compared to the alternative presented here.

In this paper MCMC techniques are used to obtain fully Bayesian reconstructions. Here, the parameters in the prior are themselves considered drawn from appropriate hyperpriors. Thus, unlike reconstructions using parameter point estimates, the uncertainty in their values is acknowledged. In Section 2 the fully Bayesian model is presented: the form of prior is described and it is shown that the expression for its normalisation constant can be analytically simplified, reducing it to an intractable function of lower dimension. The use of MCMC techniques to obtain estimates of functionals of the posterior distribution is presented in Section 3. Proposal distributions and acceptance ratios for use in the Hastings algorithm are given and practical considerations of implementation are discussed, in particular the use of reverse logistic regression as a method to estimate the normalisation constant. In Section 4 the results on simulated, physical phantom and real data are presented. The paper concludes with a brief discussion in Section 5.

2 Fully Bayesian model

The fully Bayesian model has an extra level in the hierarchy to include hyperpriors for the prior's parameters, the full joint posterior being given by

$$p(\mathbf{x}, \beta, \delta|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta, \delta)p(\beta, \delta). \tag{2.1}$$

The data model, $p(\mathbf{y}|\mathbf{x})$, is given in equation (1.1). The prior model used for $\mathbf{x}$ is

$$p(\mathbf{x}|\beta,\delta) \propto \exp\left\{-\epsilon\delta^2 \sum_s x_s^2 - \beta(1+\beta)\sum_{s\sim r} w_{sr} \log\cosh\left(\frac{\delta(x_s - x_r)}{\beta}\right)\right\}$$

where ϵ is a positive constant and w_{sr} is a weight coding the strength of neighbourliness between pixel s and r:

$$w_{sr} = \begin{cases} 1 & \text{if } s \text{ and } r \text{ are orthogonal nearest neighbours,} \\ \sqrt{1/2} & \text{if } s \text{ and } r \text{ are diagonal nearest neighbours,} \\ 0 & \text{otherwise.} \end{cases}$$

Without the term in the exponent involving ϵ, the prior would be a pairwise difference prior with a log cosh potential function. This was proposed by Green (1990) and gives a flexible family of prior distributions ranging from absolute value to Gaussian as the shape parameter $\beta \to 0$ and ∞ respectively. The parameter δ is an inverse scale parameter. We would ideally like to target this prior, but since the range of values of $\{x_s\}$ is unbounded, the prior is improper. This does not matter for Bayesian reconstructions with fixed prior parameters since it nevertheless gives rise to a proper posterior distribution. However, here we shall require propriety and thus the term involving ϵ is included; prior experimentation is used to fix it to a small value so that the contribution of the first term in the exponent is negligible compared to the second. Both of the parameters, β and δ, are restricted to be greater than zero, and, with this in mind, the prior distribution for them comprise independent gamma distributions for each parameter characterised by $\Gamma(r_\beta, s_\beta)$ and $\Gamma(r_\delta, s_\delta)$ respectively. The hyperparameter values are chosen to give non-informative hyperpriors.

Inference about $\mathbf{x}$ is based on the posterior distribution $p(\mathbf{x}|\mathbf{y})$ which arrises from integrating out β and δ from equation (2.1). This non-trivial task is suited to MCMC techniques which will be discussed in the next section. However, as we shall see in Section 3.1, this requires knowledge of the normalisation constant for $p(\mathbf{x}|\beta,\delta)$ which is an intractable function of β, δ, the fixed value of ϵ and the dimensions of the pixel grid. For a given ϵ and a particular pixel grid the problem can be analytically simplified reducing it to an intractable function of single dimension. Let

$$p(\mathbf{x}|\beta,\delta) = \frac{1}{Z(\beta,\delta)} \exp\left\{-\epsilon\delta^2 \sum_s x_s^2 - \beta(1+\beta)\sum_{s\sim r} w_{sr} \log\cosh\left(\frac{\delta(x_s - x_r)}{\beta}\right)\right\}$$

where the normalisation constant is given by

$$Z(\beta,\delta) = \int_0^\infty \exp\left\{-\epsilon\delta^2 \sum_s x_s^2 - \beta(1+\beta)\sum_{s\sim r} w_{sr} \log\cosh\left(\frac{\delta(x_s - x_r)}{\beta}\right)\right\} \mathrm{d}\mathbf{x}.$$

Making the vector substitution $\delta \mathbf{x} = \mathbf{v}$, this can be rewritten as:

$$\begin{aligned} Z(\beta, \delta) &= \frac{1}{\delta^n} \int_0^\infty \exp\left\{ -\epsilon \sum_s v_s^2 - \beta(1+\beta) \sum_{s \sim r} w_{sr} \log\cosh\left(\frac{v_s - v_r}{\beta} \right) \right\} d\mathbf{v} \\ &= \frac{Z(\beta, 1)}{\delta^n}, \end{aligned}$$

therefore the prior for $\mathbf{x}$ can be re-expressed as:

$$p(\mathbf{x}|\beta, \delta) = \frac{\delta^n}{Z(\beta, 1)} \exp\left\{ -\epsilon \delta^2 \sum_s x_s^2 - \beta(1+\beta) \sum_{s \sim r} w_{sr} \log\cosh\left(\frac{\delta(x_s - x_r)}{\beta} \right) \right\}.$$

Thus, with ϵ fixed and for a specific pixel grid, effectively we only need knowledge of the normalisation constant of the one parameter prior $p(\mathbf{x}|\beta, 1)$. Although this is intractable it can be estimated by off-line experimentation (see Section 3.2.1).

3 Markov chain Monte Carlo reconstructions

In this section we shall consider obtaining estimates of functionals of the posterior distribution $p(\mathbf{x}|\mathbf{y})$. These are obtained using MCMC methods which involve the construction of a Markov chain that converges to the distribution of interest. Thus, after discarding an initial transient, a sequence of simulations from this chain can be used as a dependent sample from which suitable statistics can be gathered. In this study the Hastings algorithm is used to construct the required chain. For a full account of MCMC methods see Besag, Green, Higdon and Mengersen (1995). Green (1995) presents MCMC reconstructions using fixed prior parameters.

3.1 Hastings algorithm

Let the current set of model parameter values be denoted by $\boldsymbol{\omega} = (\mathbf{x}, \beta, \delta)$. The parameter space is systematically swept through considering a change to individual parameters leaving the current values of the others untouched. Thus at each stage we have a proposed update, $\boldsymbol{\omega}'$, whose difference from $\boldsymbol{\omega}$ is only in one parameter. Let the probability density for changing from $\boldsymbol{\omega}$ to $\boldsymbol{\omega}'$ be given by $q(\boldsymbol{\omega}, \boldsymbol{\omega}')$, then the proposal is accepted and the parameter space updated accordingly with probability

$$\min\left\{ 1, \frac{p(\boldsymbol{\omega}'|\mathbf{y})}{p(\boldsymbol{\omega}|\mathbf{y})} \frac{q(\boldsymbol{\omega}', \boldsymbol{\omega})}{q(\boldsymbol{\omega}, \boldsymbol{\omega}')} \right\}. \tag{3.1}$$

We shall refer to the two fractions in the acceptance ratio (3.1) as the posterior and proposal ratios respectively. After each sweep through the parameter space, the resulting $\boldsymbol{\omega}$ form a time-homogeneous Markov chain that converges to the full joint posterior distribution $p(\mathbf{x}, \beta, \delta|\mathbf{y})$. The posterior expectation of a pixel's value can be estimated by averaging over a number of realisations formed by MCMC iterations at convergence, thus effectively integrating out β and δ.

The form of $q(\boldsymbol{\omega}, \boldsymbol{\omega}')$ is different depending on the parameter type being changed. These are now given for the parameter types together with the relevant acceptance probabilities.

3.1.1 x_i updates

The pixel space is systematically visited and at each visit it is proposed to change $\mathbf{x}$ to $\mathbf{x}'$ and hence $\boldsymbol{\omega}$ to $\boldsymbol{\omega}'$. At the visit to the i^{th} pixel, x_i' is drawn from a normal distribution centred at its current value that is conditioned to comply with the non-negativity constraints. All other pixels are left untouched, that is $x_j' = x_j$ for $j \neq i$. Thus the probability of changing from $\boldsymbol{\omega}$ to $\boldsymbol{\omega}'$ is given by

$$q(\boldsymbol{\omega}, \boldsymbol{\omega}') = \begin{cases} 0 & \text{if } x_i' < 0, \\ \frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{1}{2\sigma_x^2}(x_i' - x_i)^2\right) / \Phi(x_i/\sigma_x) & \text{otherwise} \end{cases}$$

where σ_x^2 is the variance of the proposal distribution. Any value may be used for the variance but it greatly influences the rate of convergence to equilibrium. This choice is considered in the Section 3.2.5. The resulting two fractions in the Hastings acceptance ratio are:

$$\begin{aligned}
\frac{q(\boldsymbol{\omega}', \boldsymbol{\omega})}{q(\boldsymbol{\omega}, \boldsymbol{\omega}')} &= \frac{\frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{1}{2\sigma_x^2}(x_i - x_i')^2\right) / \Phi(x_i'/\sigma_x)}{\frac{1}{\sqrt{2\pi\sigma_x^2}} \exp\left(-\frac{1}{2\sigma_x^2}(x_i' - x_i)^2\right) / \Phi(x_i/\sigma_x)} \\
&= \frac{\Phi\left(x_i/\sigma_x\right)}{\Phi\left(x_i'/\sigma_x\right)},
\end{aligned}$$

and

$$\begin{aligned}
\frac{p(\boldsymbol{\omega}'|\mathbf{y})}{p(\boldsymbol{\omega}|\mathbf{y})} &= \frac{p(\mathbf{y}|\mathbf{x}')p(\mathbf{x}'|\beta,\delta)p(\beta)p(\delta)}{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta,\delta)p(\beta)p(\delta)} \\
&= \frac{p(\mathbf{y}|\mathbf{x}')p(\mathbf{x}'|\beta,\delta)}{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta,\delta)} \\
&= \prod_t \frac{\exp(-\sum_s a_{ts}x_s')(\sum_s a_{ts}x_s')^{y_t}}{\exp(-\sum_s a_{ts}x_s)(\sum_s a_{ts}x_s)^{y_t}} \times \frac{\delta^n/Z(\beta,1)}{\delta^n/Z(\beta,1)} \\
&\quad \times \frac{\exp\left\{-\epsilon\delta^2 \sum_s (x_s')^2 - \beta(1+\beta)\sum_{s\sim r} w_{sr} \log\cosh\left(\frac{\delta(x_s'-x_r')}{\beta}\right)\right\}}{\exp\left\{-\epsilon\delta^2 \sum_s x_s^2 - \beta(1+\beta)\sum_{s\sim r} w_{sr} \log\cosh\left(\frac{\delta(x_s-x_r)}{\beta}\right)\right\}} \\
&= \prod_t \exp\left\{-a_{ti}(x_i' - x_i)\right\} \left(\frac{\sum_s a_{ts}x_s + a_{ti}(x_i' - x_i)}{\sum_s a_{ts}x_s}\right)^{y_t}
\end{aligned}$$

$$
\begin{aligned}
& \times \exp\left[-\epsilon\delta^2\left((x_i')^2 - x_i^2\right) \; - \beta(1+\beta)\sum_{s\sim i} w_{si}\left\{\log\cosh\left(\frac{\delta(x_i' - x_s)}{\beta}\right)\right.\right. \\
& \left.\left. - \log\cosh\left(\frac{\delta(x_i - x_s)}{\beta}\right)\right\}\right] \\
= \; & \exp\left[\sum_{\{t:a_{ti}>0\}}\left\{y_t \log\left(1 + \frac{a_{ti}(x_i' - x_i)}{\sum_s a_{ts}x_s}\right) - a_{ti}(x_i' - x_i)\right\}\right. \\
& + \epsilon\delta^2\left(x_i^2 - (x_i')^2\right) + \beta(1+\beta)\sum_{s\sim i} w_{si}\left\{\log\cosh\left(\frac{\delta(x_i - x_s)}{\beta}\right)\right. \\
& \left.\left. - \log\cosh\left(\frac{\delta(x_i' - x_s)}{\beta}\right)\right\}\right]. \qquad (3.2)
\end{aligned}
$$

Using the geometric/physical weights of Weir and Green (1993) on a 48×48 pixel grid from 64 projections of 64 counts, only about 10% of the weights $\{a_{ts}\}$ are non-zero. These are stored in a sparse matrix representation: the non-zero weights are stored columnwise together with pointers to the detectors t concerned. This allows the posterior ratio in the Hastings acceptance probability, (3.2), to be calculated very cheaply.

3.1.2 δ updates

The proposal distribution for δ' is a normal distribution centred at the current value δ that is conditioned to comply with the non-negativity constraint. Thus the proposal ratio is

$$
\frac{q(\boldsymbol{\omega}', \boldsymbol{\omega})}{q(\boldsymbol{\omega}, \boldsymbol{\omega}')} = \frac{\Phi(\delta/\sigma_\delta)}{\Phi(\delta'/\sigma_\delta)}
$$

where σ_δ is the standard deviation of the proposal distribution. The posterior ratio is

$$
\begin{aligned}
\frac{p(\boldsymbol{\omega}'|y)}{p(\boldsymbol{\omega}|y)} &= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta,\delta')p(\beta)p(\delta')}{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta,\delta)p(\beta)p(\delta)} \\
&= \frac{p(\mathbf{x}|\beta,\delta')p(\delta')}{p(\mathbf{x}|\beta,\delta)p(\delta)} \\
&= \frac{(\delta')^n/Z(\beta,1)}{\delta^n/Z(\beta,1)} \\
&\quad \times \frac{\exp\left\{-\epsilon(\delta')^2\sum_s x_s^2 - \beta(1+\beta)\sum_{s\sim r} w_{sr}\log\cosh\left(\frac{\delta'(x_s - x_r)}{\beta}\right)\right\}}{\exp\left\{-\epsilon\delta^2\sum_s x_s^2 - \beta(1+\beta)\sum_{s\sim r} w_{sr}\log\cosh\left(\frac{\delta(x_s - x_r)}{\beta}\right)\right\}} \\
&\quad \times \frac{\{(s_\delta)^{r_\delta}/\Gamma(r_\delta)\}\,(\delta')^{(r_\delta - 1)}\exp(-s_\delta\delta')}{\{(s_\delta)^{r_\delta}/\Gamma(r_\delta)\}\,\delta^{(r_\delta - 1)}\exp(-s_\delta\delta)}
\end{aligned}
$$

$$
\begin{aligned}
= &\left(\frac{\delta'}{\delta}\right)^{(n+r_\delta-1)} \times \exp\left[\epsilon\left(\delta^2-(\delta')^2\right)\sum_s x_s^2\right. \\
&+\beta(1+\beta)\sum_{s\sim r} w_{sr}\left\{\log\cosh\left(\frac{\delta(x_s-x_r)}{\beta}\right)-\log\cosh\left(\frac{\delta'(x_s-x_r)}{\beta}\right)\right\} \\
&\left.+s_\delta(\delta-\delta')\right].
\end{aligned}
$$

3.1.3 β updates

The proposal distribution for β' is a normal distribution centred at the current value β that is conditioned to comply with the non-negativity constraint. Thus the proposal ratio is

$$
\frac{q(\boldsymbol{\omega}', \boldsymbol{\omega})}{q(\boldsymbol{\omega}, \boldsymbol{\omega}')} = \frac{\Phi(\beta/\sigma_\beta)}{\Phi(\beta'/\sigma_\beta)}
$$

where σ_β is the standard deviation of the proposal distribution. The posterior ratio is

$$
\begin{aligned}
\frac{p(\boldsymbol{\omega}'|\mathbf{y})}{p(\boldsymbol{\omega}|\mathbf{y})} &= \frac{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta',\delta)p(\beta')p(\delta)}{p(\mathbf{y}|\mathbf{x})p(\mathbf{x}|\beta,\delta)p(\beta)p(\delta)} \\
&= \frac{p(\mathbf{x}|\beta',\delta)p(\beta')}{p(\mathbf{x}|\beta,\delta)p(\beta)} \\
&= \frac{\delta^n/Z(\beta',1)}{\delta^n/Z(\beta,1)} \\
&\quad \times \frac{\exp\left\{-\epsilon\delta^2\sum_s x_s^2-\beta'(1+\beta')\sum_{s\sim r} w_{sr}\log\cosh\left(\frac{\delta(x_s-x_r)}{\beta'}\right)\right\}}{\exp\left\{-\epsilon\delta^2\sum_s x_s^2-\beta(1+\beta)\sum_{s\sim r} w_{sr}\log\cosh\left(\frac{\delta(x_s-x_r)}{\beta}\right)\right\}} \\
&\quad \times \frac{\{(s_\beta)^{r_\beta}/\Gamma(r_\beta)\}\,(\beta')^{(r_\beta-1)}\exp(-s_\beta\beta')}{\{(s_\beta)^{r_\beta}/\Gamma(r_\beta)\}\,\beta^{(r_\beta-1)}\exp(-s_\beta\beta)} \\
&= \frac{Z(\beta,1)}{Z(\beta',1)}\left(\frac{\beta'}{\beta}\right)^{(r_\beta-1)} \times \exp\left[\sum_{s\sim r} w_{sr}\left\{\beta(1+\beta)\log\cosh\left(\frac{\delta(x_s-x_r)}{\beta}\right)\right.\right. \\
&\quad \left.\left.-\beta'(1+\beta')\log\cosh\left(\frac{\delta(x_s-x_r)}{\beta'}\right)\right\}+s_\beta(\beta-\beta')\right].
\end{aligned}
$$

From this we can see that we require the ratios of the intractable normalisation constants $Z(\beta,1)$ and $Z(\beta',1)$. We shall address their estimation in Section 3.2.1.

3.2 Practical considerations

Successful and efficient MCMC use requires careful consideration of many practical details. Here, we consider those relevant to this problem.

3.2.1 Estimation of the prior normalisation constant

We have seen that the MCMC method requires the ratio of normalisation constants for various parameter values. This can be estimated using reverse logistic regression which was proposed by Geyer (1994) to estimate, up to a single constant of proportionality, normalising constants from unnormalised densities that are on the same state space. Convergence and asymptotic normality of the estimates are proved under very weak regularity conditions. Here, we shall not give detailed explanations of these or the rationale behind the method but simply give the regression details applicable to this problem.

We require simulations from the unnormalised one parameter prior, $p(\mathbf{x}|\beta, 1)$, using various β values. This can be achieved using the Hastings algorithm and proposal distribution in Section 3.1.1. Thus, the pixel space is systematically visited and the proposal for x_i is accepted with probability given by

$$\min\left\{1, \exp\left[\epsilon\left(x_i^2 - (x_i')^2\right) + \beta(1+\beta)\sum_{s\sim i} w_{si}\left\{\log\cosh\left(\frac{x_i - x_s}{\beta}\right)\right.\right.\right.$$
$$\left.\left.\left. - \log\cosh\left(\frac{x_i' - x_s}{\beta}\right)\right\}\right]\frac{\Phi(x_i/\sigma_x)}{\Phi(x_i'/\sigma_x)}\right\}.$$

We need to simulate m different samples using various β values over a relevant range. Equally spaced values are chosen, for normalisation constants of values in between these interpolation will be used. Let β_j, $j = 1, \ldots, m$, be the prior values that we simulate samples from and $h_j(\mathbf{x})$ be the corresponding unnormalised one parameter priors, that is

$$h_j(\mathbf{x}) \quad = \quad \exp\left\{-\epsilon\sum_s x_s^2 - \beta_j(1+\beta_j)\sum_{s\sim r} w_{sr}\log\cosh\left(\frac{x_s - x_r}{\beta_j}\right)\right\}.$$

Furthermore, for the sample using β_j let c_j be the unknown normalising constant, n_j be the length of the converged MCMC sample, and $\mathbf{X}_{ij}$ denote the the ith MCMC sample. Let

$$p_j(\mathbf{x}, \eta) \quad = \quad \frac{h_j(\mathbf{x})e^{\eta_j}}{\sum_{k=1}^m h_k(\mathbf{x})e^{\eta_k}}$$

where

$$\eta_j = -\log c_j + \log\frac{n_j}{|n|},$$

then reverse logistic regression involves estimating the η's, which determine the c's, by maximising

$$l_n(\eta) = \sum_{j=1}^m\sum_{i=1}^{n_j}\log p_j(\mathbf{X}_{ij}, \eta).$$

Note that for each sample, $\mathbf{X}_{ij}$, the unnormalised density value $h_j(\mathbf{X}_{ij})$ needs to be calculated and saved *for all* β_j.

The log likelihood is concave and so can be maximised by many iterative schemes. Here the NAG modified-Newton algorithm E04LAF is used. This requires the following first and second derivatives:

$$\begin{aligned}
\frac{\partial l_n(\eta)}{\partial \eta_r} &= n_r - \sum_{j=1}^{m}\sum_{i=1}^{n_j} p_r(\mathbf{X}_{ij},\eta) \\
\frac{\partial^2 l_n(\eta)}{\partial^2 \eta_r} &= -\sum_{j=1}^{m}\sum_{i=1}^{n_j} p_r(\mathbf{X}_{ij},\eta)\{1 - p_r(\mathbf{X}_{ij},\eta)\} \\
\frac{\partial^2 l_n(\eta)}{\partial \eta_r \partial \eta_s} &= \sum_{j=1}^{m}\sum_{i=1}^{n_j} p_r(\mathbf{X}_{ij},\eta) p_s(\mathbf{X}_{ij},\eta), \quad r \neq s.
\end{aligned}$$

The algorithm was found to work well with initial values simply set to zero.

3.2.2 Initial values

Initial values for $\mathbf{x}$, β and δ were not found to be important providing the proposal distribution variance parameters were carefully chosen (see Section 3.2.5). A uniform start for $\mathbf{x}$, that is all x_i assigned an equal value, was adopted from deterministic algorithm use. Experience using the method has shown that the smoothing parameter, β, is always of the same approximate magnitude, and therefore an initial value can easily be assigned. The scale parameter, δ, is naturally more variable but, when the other parameters are set as above, it is not found to be sensitive to its initial value.

3.2.3 Detection of convergence

The question of when convergence is obtained needs to be addressed. A simple approach is to monitor the values of several one-dimensional functionals of the evolving process. These create stochastic processes that can be plotted against sweep number. When they appear stable, the Markov chain is assumed to have converged. For this study, we monitor the relative frequency of acceptance of pixels per sweep, several pixel estimates, the hyperparameters β and δ, the norm of the increment in body space between successive sweeps and the total posterior energy given by

$$\epsilon\delta^2 \sum_s x_s^2 + \beta(1+\beta)\sum_{s\sim r} w_{sr} \log\cosh\left(\frac{\delta(x_s - x_r)}{\beta}\right). \tag{3.3}$$

3.2.4 Chain length

The number of sweeps through the data used in the calculation of our posterior estimates needs to be considered. Our estimators are averages of functionals of the chain over time:

$$\hat{f}_N = \frac{1}{N}\sum_{t=1}^{N} f\left(\mathbf{x}^{(t)}\right)$$

then the variance of an estimator is given by

$$\begin{aligned}
\operatorname{var}\left(\hat{f}_N\right) &= \frac{1}{N^2}\sum_{s=1}^{N}\sum_{t=1}^{N}\operatorname{cov}\left\{f\left(\mathbf{x}^{(s)}\right), f\left(\mathbf{x}^{(t)}\right)\right\} \\
&\sim \frac{\sigma^2}{N}\sum_{t=-(N-1)}^{N-1}\left(1-\frac{|t|}{N}\right)\rho_f(t) \\
&\sim \frac{\sigma^2}{N}\sum_{t=-\infty}^{\infty}\rho_f(t) \qquad (3.4)
\end{aligned}$$

where $\rho_f(t)$ is the autocorrelation of the process $\{f(\mathbf{x}^{(t)})\}$ calculated at equilibrium and σ^2 is the equilibrium variance of $f(\mathbf{x})$. The dependence within the chain due to the Markovian nature of the algorithm thus leads to the asymptotic variance being a factor of

$$\tau(f) = \sum_{t=-\infty}^{\infty}\rho_f(t)$$

times greater than that which would have been obtained from independent sampling. This factor is called the *integrated autocorrelation time*. Aykroyd and Green (1991) aim for such 'Monte Carlo variances' which do not exceed 1% of the sampling variances of the estimators. We shall use this recommendation to determine the length of our sequence. We therefore require the length of our chain, N, to satisfy

$$\frac{1}{N}\sum_{t=-\infty}^{\infty}\rho_f(t) < 0.01 \qquad (3.5)$$

and thus need to estimate the integrated autocorrelation time. We shall do this using the truncated periodogram method suggested by Sokal (1989):

$$\hat{\tau}(f) = \sum_{|t|\leq M}\hat{\rho}_t(f)$$

where the window width M is chosen adaptively as the minimum integer with

$$M \geq 3\hat{\tau}(f).$$

If after a run the condition (3.5) has not been met, then the estimated integrated autocorrelation times can be used to decide the number of extra sweeps through the data that are needed. We shall check this condition using the monitored pixel and prior parameters values.

3.2.5 Proposal distribution parameters

The values of the parameters in the proposal distributions will influence the rate at which the chain converges to equilibrium. Besag *et al* (1995) recommend giving the proposal distributions a spread somewhat larger than that typical of the underlying conditional distribution. In order to approximate them pilot MCMC runs created with arbitrary proposal variances can be used; after detecting and discarding the transients, the average of the respective sampling variances of the individual parameters can be used as approximations. Alternative approaches include choosing the value that minimises the estimated integrated autocorrelation time from several pilot runs or from monitoring the acceptance rate. With regard to the latter, Besag *et al* (1995) suggest that a rate between 30% and 70% for each variable often produces satisfactory results. Here, we take heed of all of the above approaches.

4 Results

In this section results for the estimation of the normalising constants are presented. These are then used to obtain fully Bayesian reconstructions from simulated, physical phantom and real data. These are compared to those obtained by *filtered back-projection* and the OSL solution to the posterior mode of the Bayesian model with fixed prior parameter values chosen from training set data. The former is the standard reconstruction technique in use today and does not use a statistical model but rather operates entirely linearly on the data. For a detailed account see Webb (1988), Chapter 4. The fully Bayesian reconstructions comprise pixelwise means, and thus some care needs to be taken when interpreting comparisons to the OSL Bayesian mode solution. Real data and filtered back-projection reconstructions were supplied by the Bristol Royal Infirmary.

4.1 Normalising constants

The work presented in the following sections requires knowledge of the normalising constants for two pixel grids of dimensions 32×32 and 48×48. For both pixel grids the value of $\epsilon = 1 \times 10^{-6}$ was used; with this value the contribution to the posterior energy (3.3) due to the term involving ϵ was never observed to be more than 0.1% of the total. From previous work using the OSL algorithm (Weir (1993)), an anticipated range of interest for MCMC β updates was identified consisting of 13 equally spaced values between 0.025 and 0.325. For each β value appropriate proposal variances were found by pilot runs and then a sample of 1000 realisations from the relevant prior were simulated. For all of the β values this length was found to be adequate using the criteria (3.5). The practical details are omitted here as an example of the methodology is given in the next section. The resulting reverse logistic regression estimates are displayed in Figure 1 (a).

It can be seen that the estimates strongly suggest an exponential relationship. However, rather than fit a parametric curve to the data points, cubic splines are used to interpolate between calculated values.

Initial experimentation with the MCMC algorithm revealed that the anticipated range was not useful as β updates fell below the smallest range value. The procedure was successively repeated until a suitable range that did not require any extrapolation was found. The final values consist of 20 evenly spaced points from 1×10^{-6} to 1.901×10^{-2}, the resulting reverse logistic regression estimates are displayed in Figure 1 (b). The same range was found to be suitable for the 48×48 pixel grid; the regression estimates are displayed in Figure 1 (c).

A simple check on the estimates can be performed using the simulated tempering algorithm of Geyer and Thompson (1993). The algorithm randomly steps through the ordered β values generating realisations from the various prior distributions. If we let $\beta_1, \ldots, \beta_i, \ldots, \beta_m$ be m ordered β values, the Hastings algorithm updates $\mathbf{x}$ using the current visited unnormalised distribution, say $h_i(\mathbf{x})$. It is then proposed to visit a different unnormalised distribution with the following probability:

$$q(i,j) = \begin{cases} 1 & \text{for } q(1,2) \text{ and } q(m,m-1), \\ 1/2 & \text{for } q(i,i \pm 1) \quad 1 < i < m, \\ 0 & \text{otherwise.} \end{cases}$$

This proposal is accepted using the Hastings ratio:

$$\min\left\{1, \frac{h_j(\mathbf{x})\pi(j)}{h_i(\mathbf{x})\pi(i)} \frac{q(j,i)}{q(i,j)}\right\}$$

where the π are called pseudo-priors to be defined later. The process generates a Markov chain of the joint distribution $(\mathbf{X}, J)$ where J is the random variable which indicates the current updating distribution for $\mathbf{x}$. The marginal distribution of J is

$$P(J = j) \propto \pi(j) \int h_j(\mathbf{x}) \mathrm{d}\mathbf{x} = c_j \pi(j)$$

where c_j is the normalising constant for distribution j. Thus, if the pseudo-priors are set to be the inverse of the corresponding reverse logistic regression estimates, the marginal distribution of J should be approximately uniform. The results for 20,000 sweeps of the algorithm for the two pixel grids are given in Tables 1 and 2. From these it can be seen that the chain mixes fairly evenly between distributions, and thus conclude that are estimates are reasonable. Note that these estimates are derived solely from the prior and are thus not peculiar to the SPECT problem.

4.2 Simulated data

The use of simulated data from known mathematical phantoms is advantageous when developing reconstruction techniques since performance can be easily as-

β_j	1×10^{-6}	1.001×10^{-3}	2.001×10^{-3}	3.001×10^{-3}	4.001×10^{-3}
η_j	24.279	22.996	21.711	20.423	19.135
count	970	972	961	994	1057
β_j	5.001×10^{-3}	6.001×10^{-3}	7.001×10^{-3}	8.001×10^{-3}	9.001×10^{-3}
η_j	17.848	16.563	15.279	13.998	12.719
count	1057	993	1011	1110	1115
β_j	1.001×10^{-2}	1.101×10^{-2}	1.201×10^{-2}	1.301×10^{-2}	1.401×10^{-2}
η_j	11.440	10.164	8.888	7.614	6.342
count	1033	1000	998	982	981
β	1.501×10^{-2}	1.601×10^{-2}	1.701×10^{-2}	1.801×10^{-2}	1.901×10^{-2}
η_j	5.070	3.800	2.531	1.265	-3.784×10^{-4}
count	968	940	937	965	956

Table 1: 32×32 pixel grid: reverse logistic regression estimates and mixing counts.

β_j	1×10^{-6}	1.001×10^{-3}	2.001×10^{-3}	3.001×10^{-3}	4.001×10^{-3}
η_j	61.503	58.419	55.542	52.599	49.624
count	982	1020	1011	1044	1069
β_j	5.001×10^{-3}	6.001×10^{-3}	7.001×10^{-3}	8.001×10^{-3}	9.001×10^{-3}
η_j	46.651	43.681	40.714	37.752	34.793
count	1095	1071	1056	1066	1063
β_j	1.001×10^{-2}	1.101×10^{-2}	1.201×10^{-2}	1.301×10^{-2}	1.401×10^{-2}
η_j	31.839	28.887	25.940	22.996	20.056
count	1021	975	965	996	978
β	1.501×10^{-2}	1.601×10^{-2}	1.701×10^{-2}	1.801×10^{-2}	1.901×10^{-2}
η_j	17.112	14.179	11.249	8.3227	5.4003
count	940	936	920	922	870

Table 2: 48×48 pixel grid: reverse logistic regression estimates and mixing counts.

sessed visually and quantitatively through, for example, the root-mean-square error, that is

$$\mathrm{RMSE} = \sqrt{\frac{1}{n}\sum_s(\hat{x}_s - x_s)^2},$$

where $\hat{x}_s$ is the estimate of the true concentration, x_s, and n is the total number of pixels. It is therefore necessary to design suitable phantoms that reflect circumstances of interest. The phantom used here (Figure 2 (a)) is similar to the phantom used in several papers, notably Kaufman (1987), and is designed to imitate the brain's typical tracer concentration. It consists of eight ellipses that define regions which include an area of low metabolic activity area (corresponding to the skull) and high metabolic activity (which represent ventricles and tumours). It was created on a 640×640 grid and then reduced to a 32×32 by averaging over pixels, with the result that the discontinuities between homogeneous regions are slightly stepped between regions. This feature is not present in Kaufman's phantom and is intended to make the phantom more realistic. The concentration values in the homogeneous regions within the brain phantom are 200, 40 and 100 corresponding to the high, low and remaining regions respectively.

Simulated data representing a system with 32 detectors at 32 angles with a total of 440,040 photon counts were created from the brain phantom by Weir (1993). These dimensions are smaller than those generally used in reality, allowing faster reconstructions. The resulting data was used as a training set from which an insight to appropriate fixed values for the prior were found. In order to identify such values the RMSE at convergence using the OSL algorithm from a uniform start was taken as the measure of performance. A coarse grid of β and δ values was successively refined in the locality of the parameter combination which at each stage attained minimum RMSE. This was repeated until the extra refinement of the grid did not significantly change the current minimum RMSE value. The process revealed a ridge of parameter combinations that gave approximately the global minimum value. A combination of these was chosen that was far away from the non-convergence region of the OSL algorithm that the process identified. The chosen values were $\beta = 0.178924$ and $\delta = 2.20377 \times 10^{-2}$ and resulted in the reconstruction displayed in Figure 2 (b). From this it can be seen that all but the small features of the truth have been captured. For the data sets in the next sections we shall compare reconstructions from the fully Bayesian method to OSL ones using these parameter values.

The fully Bayesian model requires specification of the hyperparameters of the hyperpriors of β and δ. The values chosen were $r_\beta = r_\delta = 1$ and $s_\beta = s_\delta = 0.05$ which make the hyperpriors very non-informative. For the MCMC runs it was decided to monitor the estimates of three pixels from different regions of the brain phantom corresponding to high, medium and low concentration. A pilot MCMC run of 1000 sweeps with the proposal variances $\sigma_x^2 = 10$ and $\sigma_\beta^2 = \sigma_\delta^2 = 1 \times 10^{-6}$ was carried out. The monitoring statistics are displayed in Figure 3, where the

parameter	mean	$\hat{\tau}(f)$	M	$\hat{\tau}(f)$/N	s.e.
high	189.03	21.688	68	0.02711	1.7056
med	95.266	23.883	73	0.02985	1.4096
low	46.925	23.848	73	0.02981	1.9029
β	3.4758×10^{-3}	31.591	96	0.03949	5.1290×10^{-4}
δ	2.3445×10^{-2}	75.596	227	0.09450	2.6607×10^{-4}

Table 3: Brain phantom pilot MCMC run: estimates of posterior means, integrated autocorrelation times and standard errors.

graphs of the pixel values and prior parameters clearly show the high dependency of the sample. From these graphs it appears that equilibrium has been reached at around 200 iterations. The following calculations are thus based on data from the last 800 sweeps.

The three sample variances of the pixels are 107.30, 66.55 and 121.47 for the high, medium and low pixels respectively, giving an average of 98.44. From this it was decided to let the variance of the proposal distribution for the major MCMC run be 100. The sample variances for β and δ are 6.662×10^{-6} and 7.492×10^{-7} respectively; from these the variance proposals for the major MCMC were set to 1×10^{-5} and 1×10^{-6} respectively. Table 3 gives for the three pixels and prior parameters the estimates of the posterior means, integrated autocorrelation times (together with the corresponding value of M), the statistic (3.5) relating to the requirement of the chain length and standard errors (calculated from substitution of sampling variances and estimated integrated autocorrelation times into (3.4)). It can be seen that there is a high dependence and that in order to meet our Monte Carlo variance requirements, δ's estimate of $\tau(f)$ suggests that the chain length needs to be over 7560.

The second MCMC run of length 1000 with the new proposal variances resulted in monitoring statistics that reached equilibrium a lot quicker than the previous run and did not demonstrate such a high temporal dependence (see Figure 4). The chain appears to be in equilibrium after 100 iterations. In Table 4 the chain statistics calculated from the last 900 sweeps are displayed. The estimated integrated autocorrelation times suggest that the sample length needs to be longer; it was decided to run a further 400 sweeps. The integrated autocorrelation times were recalculated with the new chain length of 1300 and gave no further cause for concern.

The final realisation from the run and the resulting estimate of the posterior mean are displayed in Figures 2 (c) and (d) respectively. The fully Bayesian posterior mean is not as aesthetically pleasing as the posterior mode from the OSL using the fixed 'best' prior parameters. The former is slightly noisier and has RMSE value of 6.3670 compared to 5.7003 from the fixed parameter posterior

parameter	mean	$\hat{\tau}(f)$	M	$\hat{\tau}(f)$/N	s.e.
high	188.29	12.392	38	0.01377	1.6785
med	93.524	8.1245	25	0.00903	0.8502
low	41.905	12.653	38	0.01406	1.4670
β	3.6189×10^{-3}	6.0526	19	0.00673	2.2004×10^{-4}
δ	2.3384×10^{-2}	11.003	34	0.01223	9.6926×10^{-5}

Table 4: Brain phantom major MCMC run: estimates of posterior means, integrated autocorrelation times and standard errors.

mode. The MCMC run took 940.3 seconds on a SUN SPARC 10/41 workstation compared to the 30.0 seconds that OSL took. Therefore, the comparative computational cheapness of OSL when 'good' fixed values of the prior parameters are available discourages the use of MCMC to obtain reconstructions. However, in reality the variety of organs imaged and range of photon counts recorded, may mean that these are not available. Using other parameter values the OSL algorithm invariably produced grossly over smoothed or noisy reconstructions. Thus the automation and quality of the fully Bayesian reconstruction is admirable.

Furthermore, the fully Bayesian method also allow us to go beyond mere reconstruction since functionals of the body space can easily be obtained as a by-product of the MCMC. For instance estimates of fully Bayesian posterior pixelwise modes could be obtained. These may be useful for examining a specific region of interest; the OSL fixed parameter solution would not be capable of this as it is a mode for the whole bodyspace. Fully Bayesian posterior medians can easily be obtained; here, the reconstruction composed of pixelwise medians was found to be very similar in appearance to the posterior mean. Also, nonparametric estimates of variability leading to *credibility bounds* for the true values can be obtained via pixelwise quantiles and order statistics. Here, we simply report pixelwise quantiles, however Besag *et al* (1995) outline the construction of simultaneous bounds for the whole vector $\mathbf{x}$. In Figure 5 two transects of the brain phantom are graphed that cross distinct regions of interest. The posterior mean and 90% and 95% pixelwise credibility bounds are superimposed on these plots. It can be seen that the posterior mean faithfully reproduces the general truth shape; the major departure being over-estimation of the low metabolic activity region. The credibility bounds, however, fully bracket the truth.

We can also use output from the MCMC to get posterior density estimates of the prior parameters; various representations of these are given in Figure 6. In (a) two estimates of $\beta's$ posterior density are displayed using histogram and kernel density representations, the latter being more informative; the density appearing to be gamma shaped. In (b) the density estimates for δ suggest a normal distribution. In (c) the joint posterior density is displayed using a grey-scale image of a 2-dimensional histogram.

4.3 Physical phantom data

The physical phantom consists of a cylindrical acrylic tank into which an acrylic insert is placed before being filled with a radio-pharmaceutical. The insert consists of eight pairs of holes drilled through the block, each pair separated by a distance equal to the hole diameters. The diameter of each pair is 25% larger than that of the preceding pair. The eight pairs form a V-shape either side of which are drilled two larger holes. The 18 holes form hot spots of the radio-pharmaceutical which simulates brain and bone scans.

In Figure 7 we compare different reconstructions for the insert. In the filtered back-projection reconstruction, the larger of the hot spots are visible but these are not as well defined as in the reconstructions from the statistical models. Note the presence of spurious radial artifacts, one of the well-documented deficiencies of the method. The statistical reconstructions are clearly an improvement over the filtered back-projection reconstructions. The OSL posterior mode reconstruction only managed to distinctly identify two of the hot spots in the V-shape whereas the fully Bayesian reconstruction gives strong hints to others' existence. In figure 8 (a) and (b) the posterior means and credibility bounds of two transects are given. These are chosen to cross through hot spots and can clearly be seen. Posterior density estimates are given in (c)-(e). Again gamma and normal distributions are suggested for β and δ respectively. The joint posterior density estimate reveals a single mode. Approximate values for this mode were used to gain an OSL fixed parameter reconstruction; it was interesting to note that this reconstruction was grossly over smoothed.

4.4 Real data

The data relate to horizontal slices through the brain and pelvis. The reconstructions of the brain (Figure 9) clearly demonstrate the inadequacy of filtered back-projection. The statistical reconstructions reveal clear isotope patterns: the OSL fixed parameter solution being smoother than the fully Bayesian reconstruction which suggests greater contrast between several regions. In Figure 10 the posterior mean and credibility bands of two cross-sections are displayed, along with posterior density estimates of the prior parameters. The pelvis reconstructions (Figure 11) reveal little difference between the statistical reconstructions whilst offering a vast improvement over filtered back-projection. Posterior mean and credibility bands of two cross-sections and prior posterior density estimates are displayed in Figure 12.

5 Discussion

The work presented here has demonstrated the ability to obtain representative Bayesian reconstructions of SPECT data without the need for ad-hoc parameter selections. The process is thus fully automated and can therefore be applied with no supervision. The reconstructions are of comparable quality to OSL ones when 'good' fixed prior parameter values are used and better when inappropriate values are used. Since MCMC methods are used, useful diagnostic by-products can easily be obtained with little further expense. The reconstructions are in comparison computationally demanding, but the time wasted in off-line prior parameter estimation needs to be taken into account. The fully Bayesian reconstruction did need its own off-line estimation for the prior normalisation constant. This was found to be straight forward using reverse logistic regression and unlike the estimation for the fixed parameters is independent of the object being imaged.

However, it is clear that under the present SPECT set up using the log-cosh pairwise Gibbs prior, fine features are missed. It therefore seems necessary to change the prior in some way, especially as the advent of new MCMC methodology allows us to be more adventurous in our modelling. It is the intention in further work to develop a prior that allows a smoothing parameter to vary stochastically across the scene.

6 Acknowledgements

I would like to thank Peter Green for stimulating discussions and Bristol Royal Infirmary for providing the data.

References

Aykroyd, R.G. and Green, P.J. (1991) Global and local priors, and the location of lesions using gamma-camera imagery. *Phil. Trans. R. Soc. Lond.*, A, 337, 323-342.

Besag, J.E. (1986) On the statistical analysis of dirty pictures (with discussion). *J. Roy. Statist. Soc.*, B, 48, 259-302.

Besag, J.E., Green, P.J., Higdon, D. and Mengersen, K. (1995) Bayesian computation and stochastic systems (with discussion). To appear in *Statistical Science*.

Dempster, A.P., Laird, N.M. and Rubin D.B. (1977) Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Statist. Soc.*, B, 39, 1-38.

Geman, S. and Geman, D. (1984) Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Trans. Pattern Anal. Machine Intell.*, 6, 721-741.

Geman, S. and McClure, D.E. (1985) Bayesian image analysis: an application to single photon emission tomography. *Proc. Amer. Statist. Assoc., Stat. Comp.* Sect., 12-18.

Geman, S. and McClure, D.E. (1987) Statistical methods for tomographic image reconstruction. *Bull. Int. Statist. Inst.* 52 (4), 5-21.

Geman, S., Manbeck, K.M. and McClure, D.E. (1991) Design and implementation of a statistical method for emission computed tomography. Reports in pattern analysis, Brown University, 154.

Geyer, C. J. (1994) Estimating Normalizing Constants and Reweighting Mixtures in Markov Chain Monte Carlo. Technical Report No. 568. School of Statistics, University of Minnesota.

Geyer, C. J. and Thompson, E. A. (1993). Annealing Markov chain Monte Carlo with applications to ancestral inference. Technical Report No. 589. School of Statistics, University of Minnesota. (to *J. Am. Statist. Assoc.*, revised and resubmitted 1994).

Green, P.J. (1990) Bayesian reconstructions from emission tomography data using a modified EM algorithm. *IEEE Trans. Medical Imaging*, 9, 84-93.

Green, P.J. (1995) Markov chain Monte Carlo in image analysis. *Markov chain Monte Carlo in practice*, (Gilks, W.R., Richardson, S., and Spiegelhalter, D.J., eds.), to be published by Chapman and Hall.

Kaufman, L. (1987) Implementing and accelerating the EM algorithm for positron emission tomography. *IEEE Trans. Medical Imaging*, 6, 37-51.

Shepp, L.A. and Vardi, Y. (1982) Maximum likelihood reconstruction in positron emission tomography. *IEEE Trans. Medical Imaging*, 1, 113-122.

Sokal, A.D. (1989) Monte Carlo methods in statistical mechanics: foundations and new algorithms. *Cours de Troisième Cycle de la Physique en Suisse Romande*, Lausanne.

Vardi, Y., Shepp, L.A. and Kaufman, L. (1985) A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.*, 80, 8-37.

Webb, S. (1988) *The physics of medical imaging*, Adam Hilger, Bristol and Philadelphia.

Weir, I.S. (1993) Statistical modelling and reconstructions in single photon emission computed tomography. *Ph.D thesis*, Bristol University.

Weir, I.S. and Green, P.J. (1994) Modelling data from single photon emission computed tomography. *Statistics and Images*, (Mardia, K.V. ed.), 313-338, Carfax Publishing Co., Abingdon.

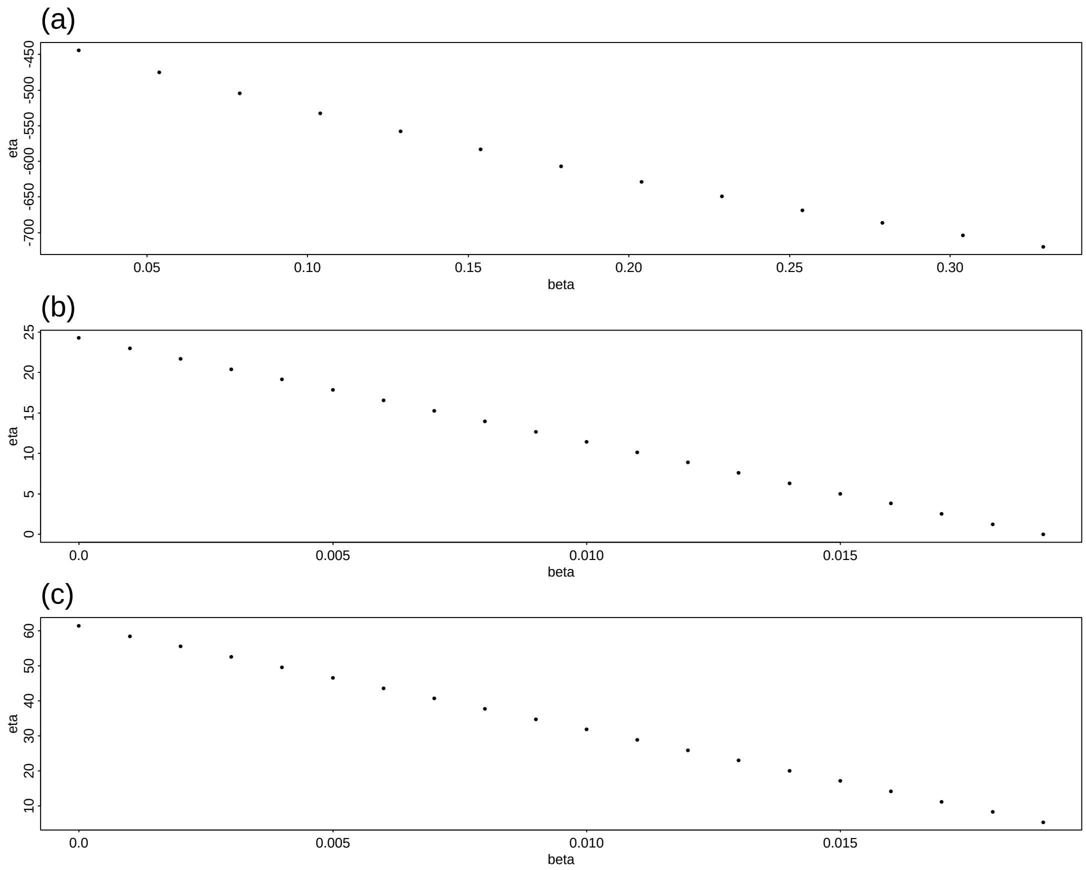

Figure 1: Reverse logistic regression estimated values: (a) 32×32 pixels, initial regression; (b) 32×32 pixels, final regression; (c) 48×48 pixels.

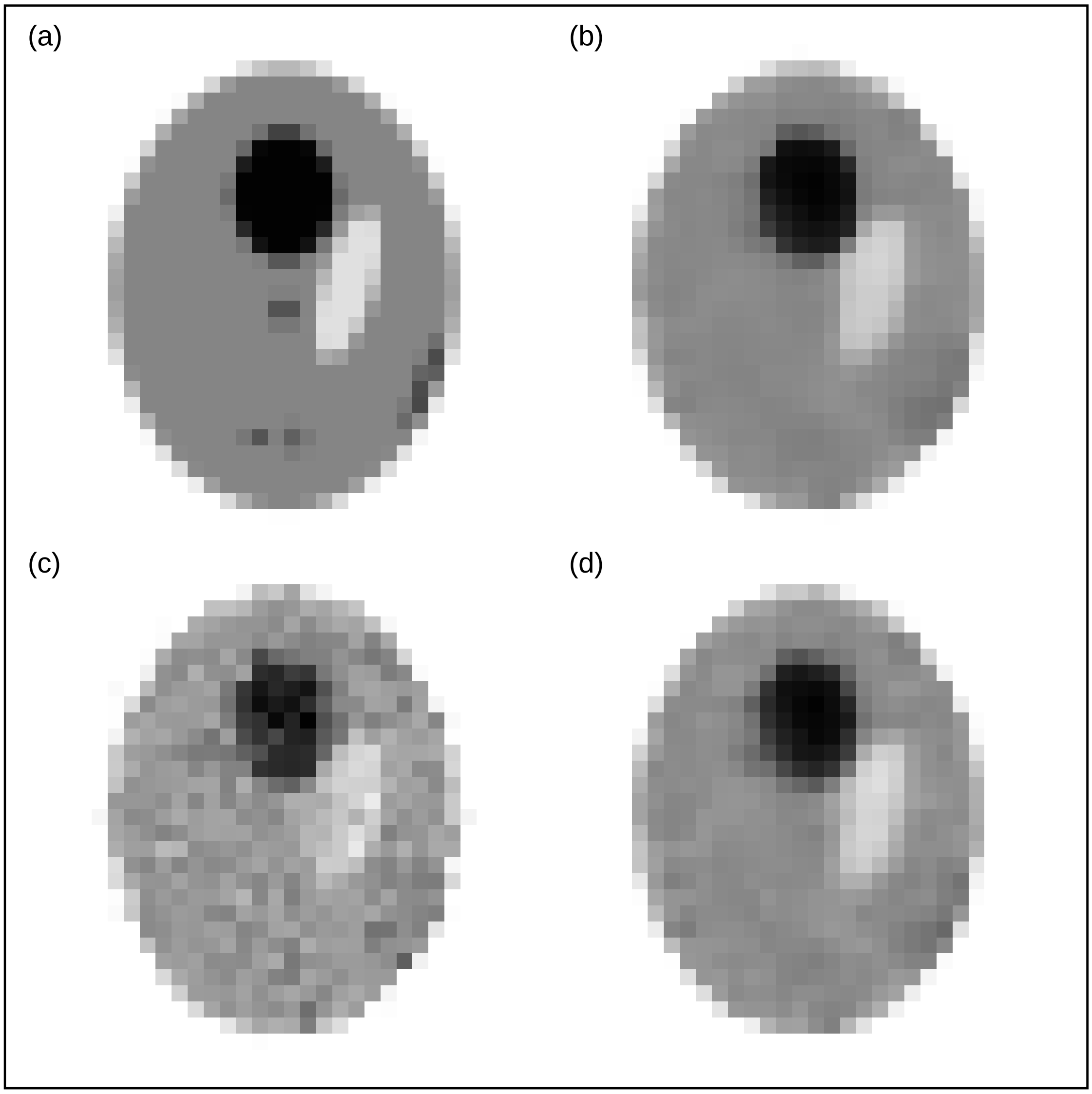

Figure 2: Brain phantom: (a) truth; (b) OSL posterior mode (fixed parameters); (c) realisation from fully Bayesian posterior; (d) fully Bayesian posterior mean.

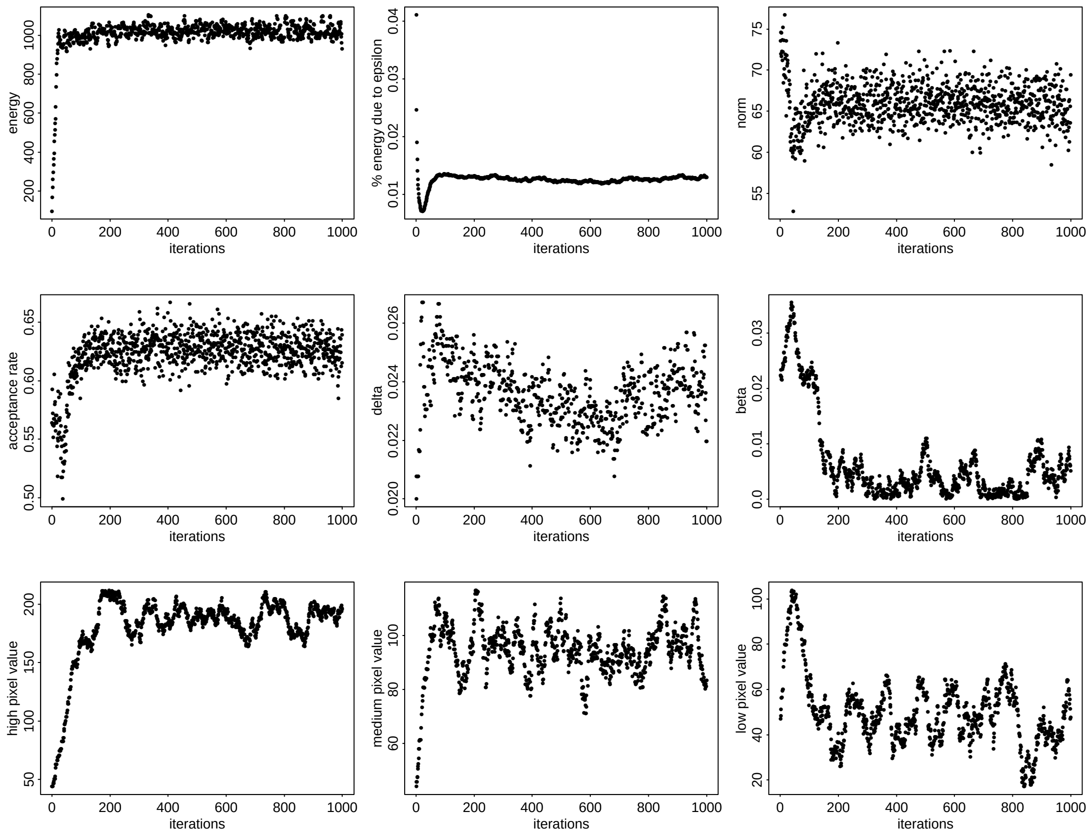

Figure 3: Brain phantom: pilot MCMC run monitoring statistics.

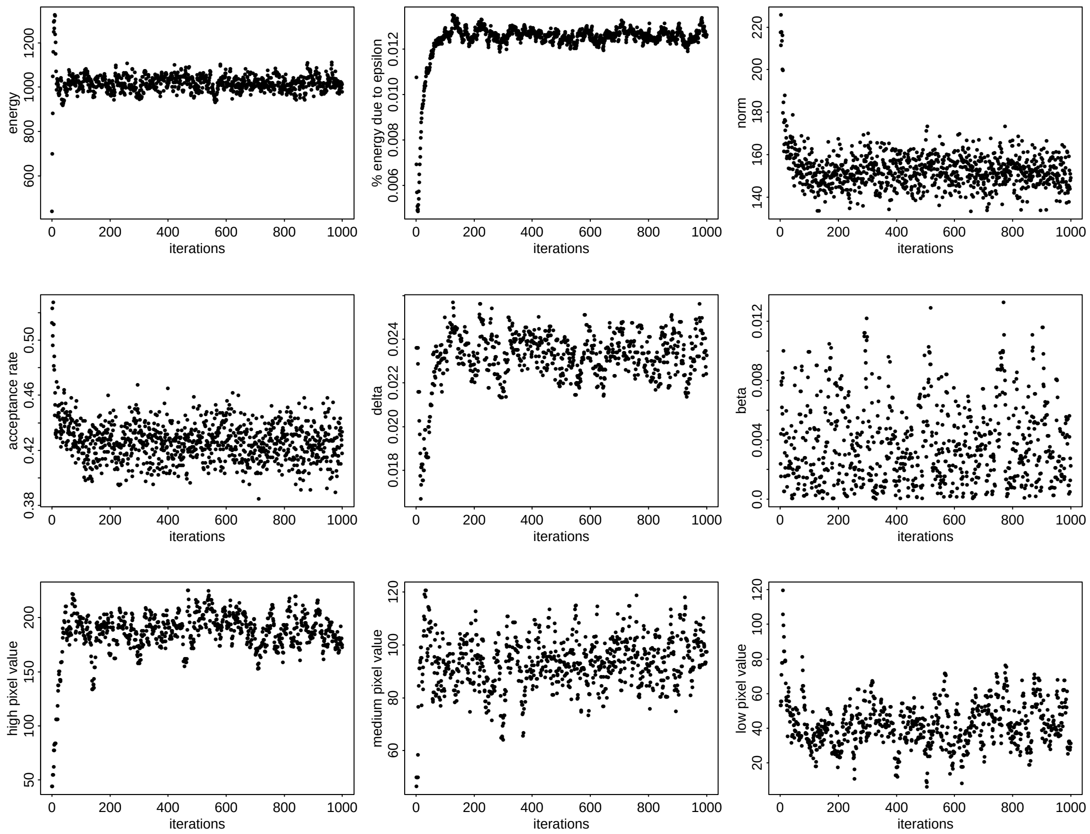

Figure 4: Brain phantom: major MCMC run monitoring statistics.

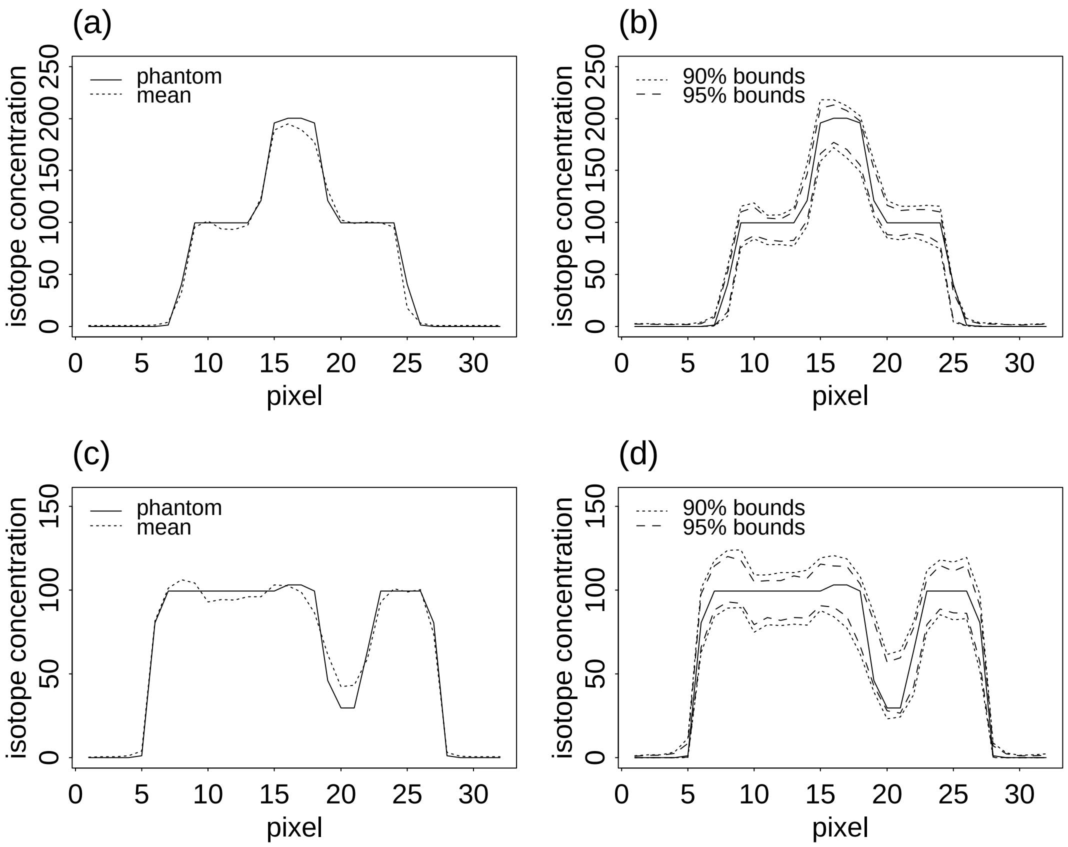

Figure 5: Brain phantom: (a) row 25 truth and posterior mean; (b) row 25 credibility bounds; (c) row 16 truth and posterior mean; (d) row 16 credibility bounds.

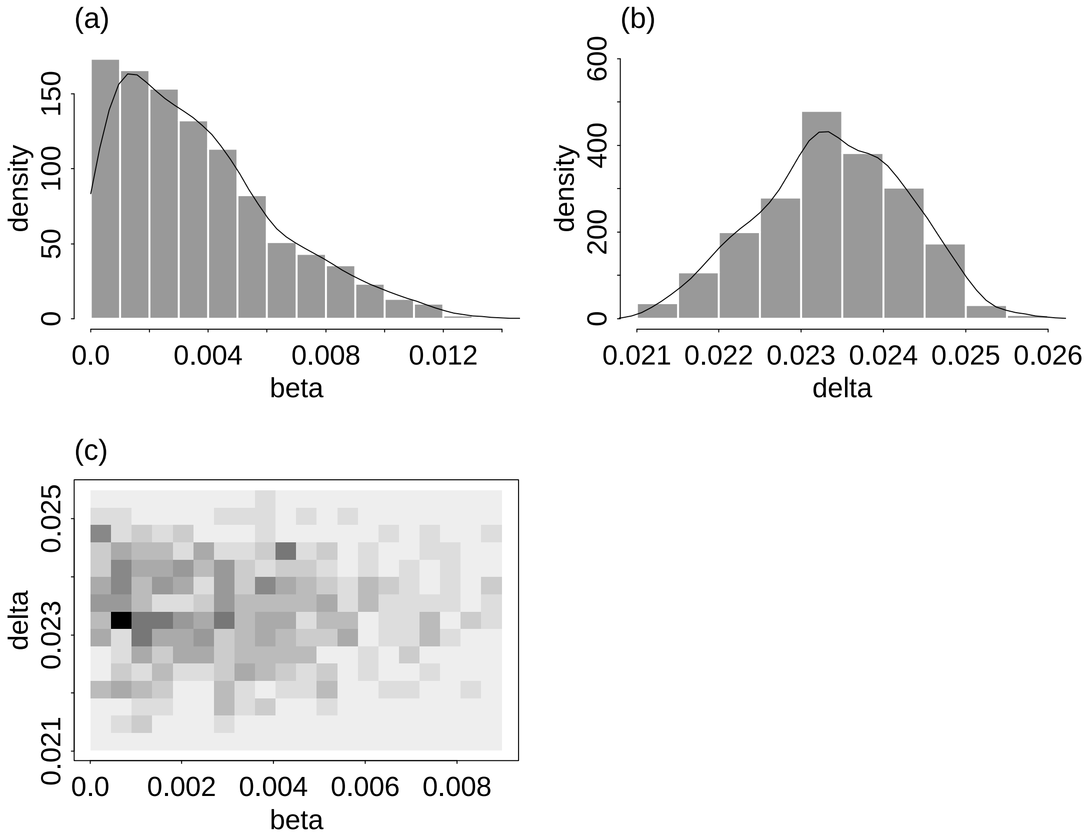

Figure 6: Brain phantom posterior density estimates: (a) histogram and kernel density representations of β's posterior density; (b) histogram and kernel density representations of δ's posterior density; (c) grey-scale 2-dimensional histogram representation of the joint posterior density of β and δ.

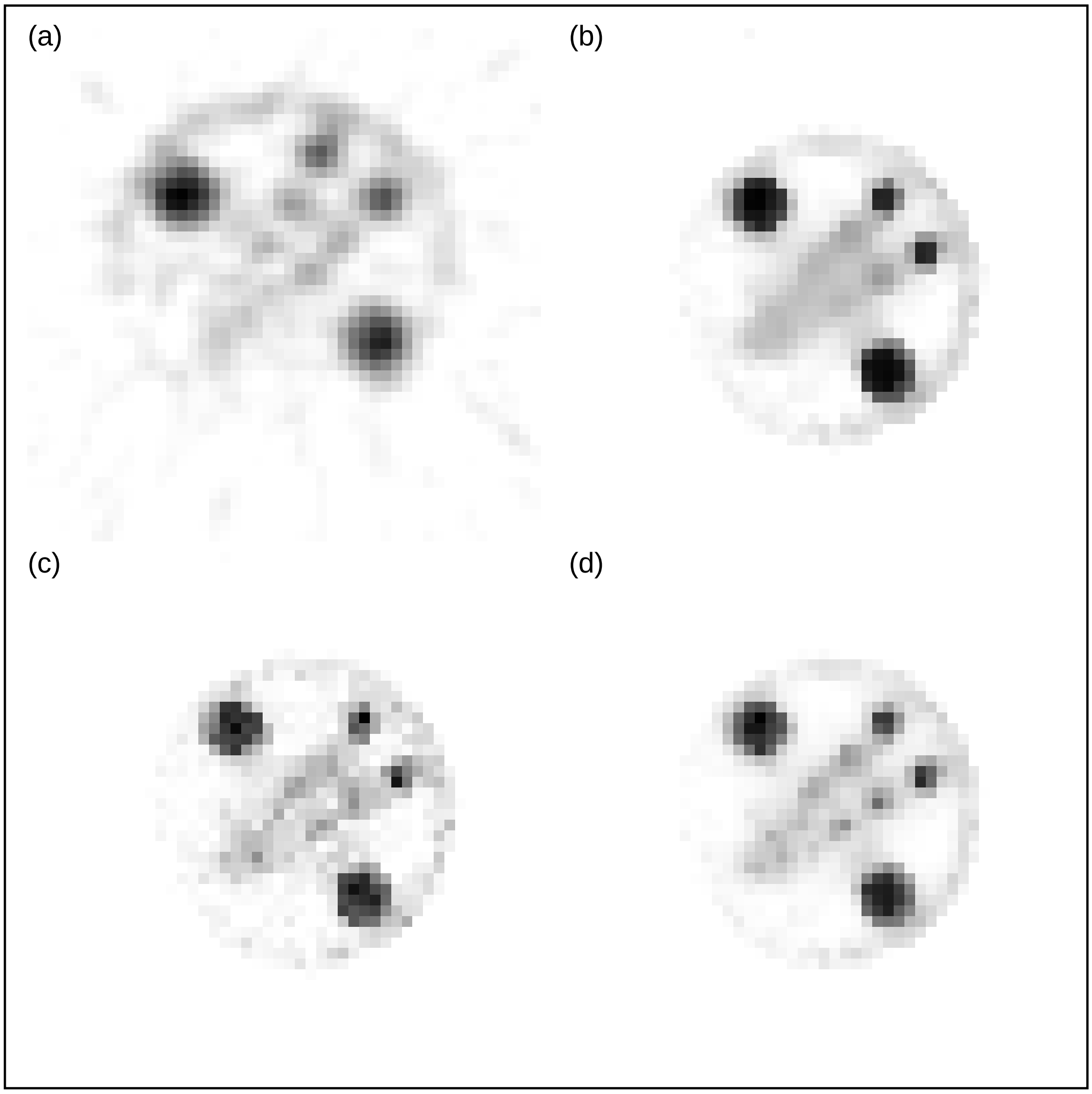

Figure 7: Reconstructions of physical phantom: (a) filtered back-projection; (b) OSL posterior mode (fixed parameters); (c) realisation from fully Bayesian posterior; (d) fully Bayesian posterior mean.

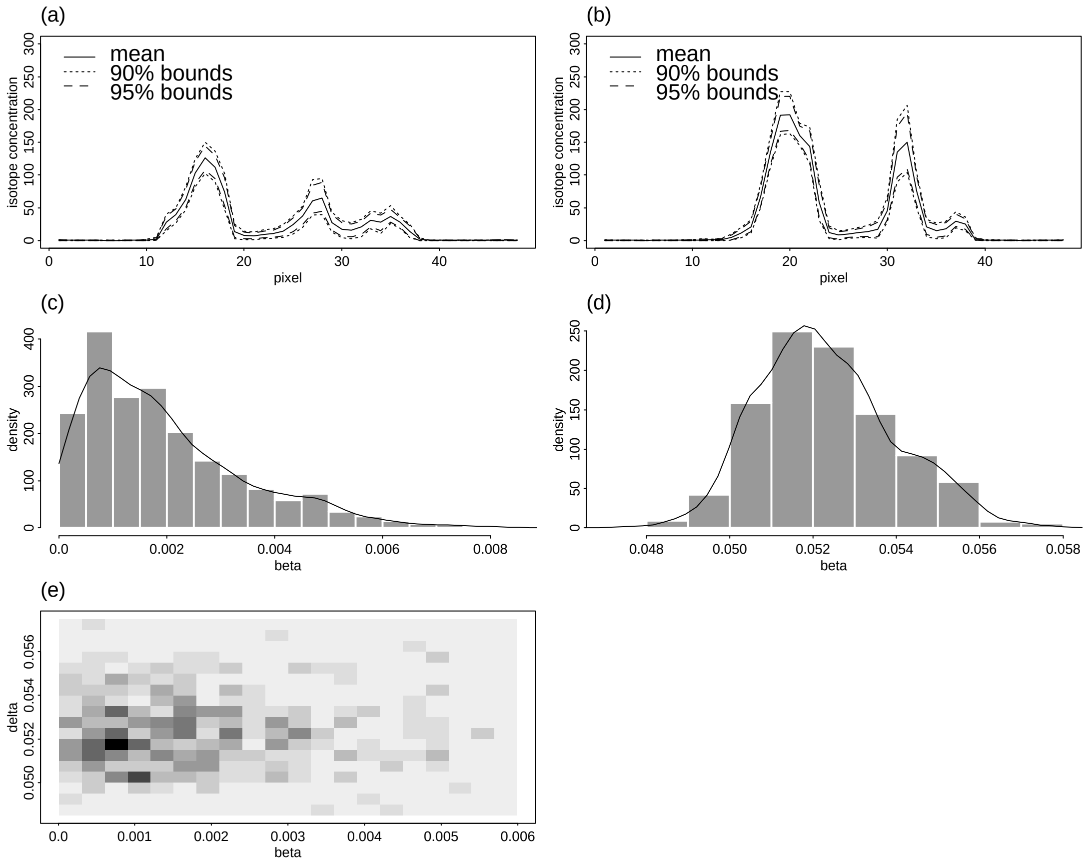

Figure 8: Physical phantom: (a) column 34 posterior mean and credibility bounds; (b) row 32 posterior mean and credibility bounds; (c) posterior density of β; (d) posterior density of δ; (e) joint posterior density of β and δ.

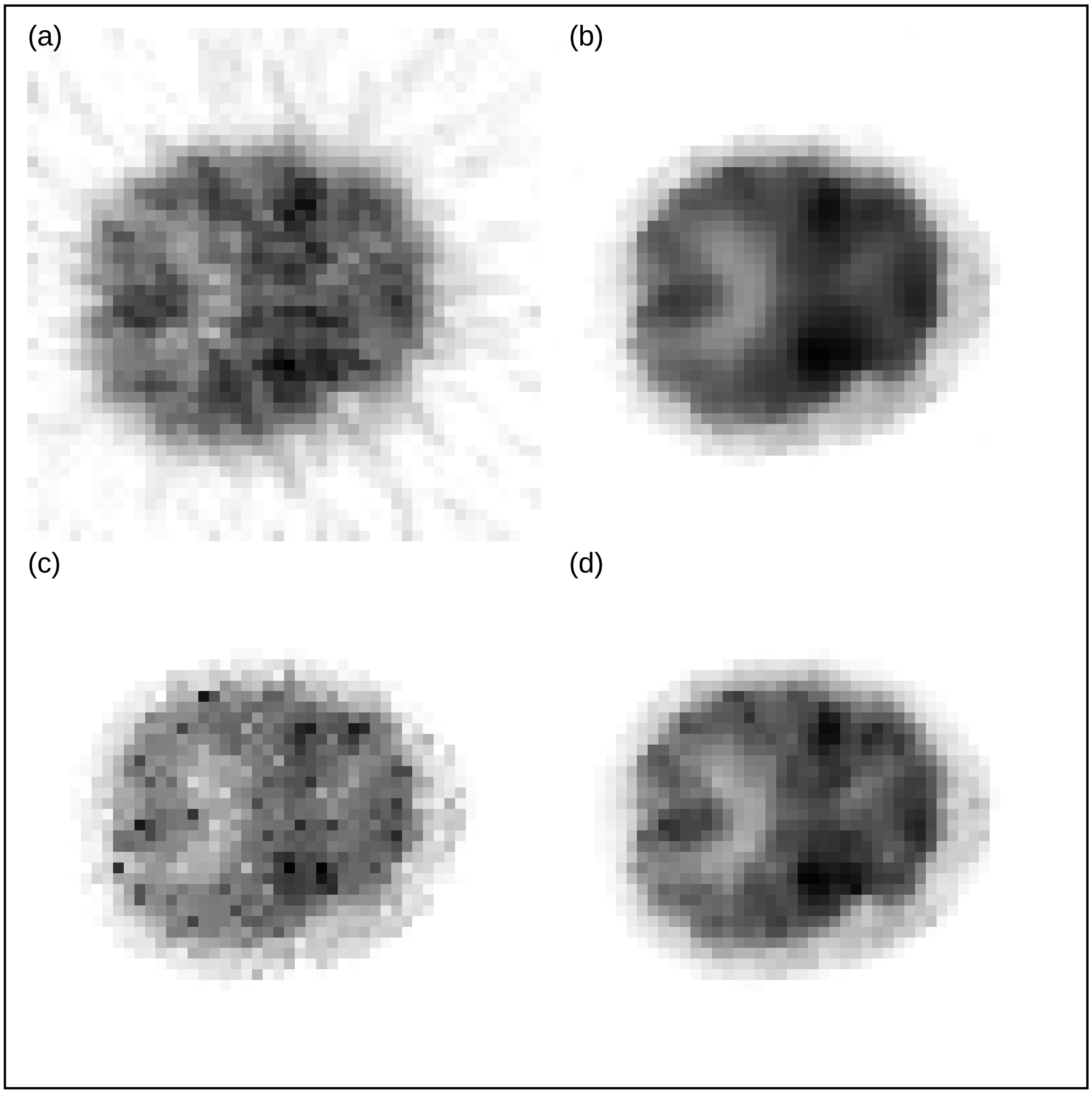

Figure 9: Reconstructions of brain: (a) filtered back-projection; (b) OSL posterior mode (fixed parameters); (c) realisation from fully Bayesian posterior; (d) fully Bayesian posterior mean.

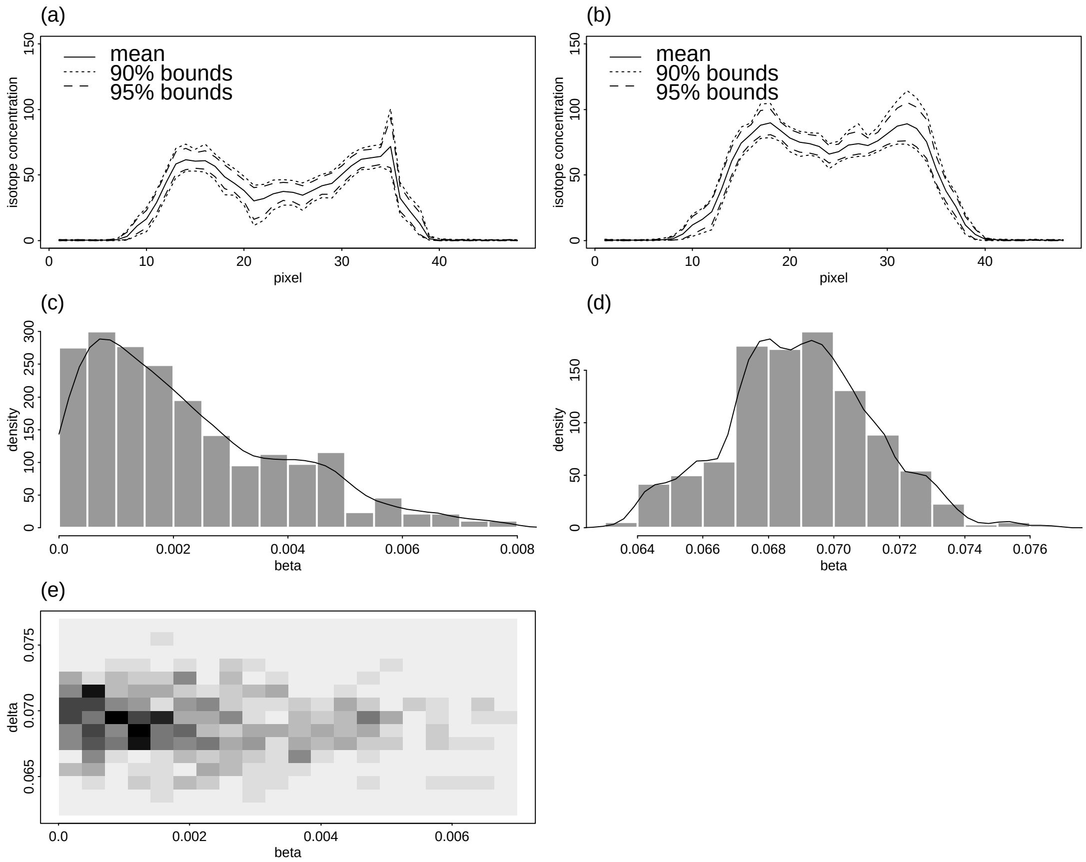

Figure 10: Brain: (a) column 20 posterior mean and credibility bounds; (b) column 26 posterior mean and credibility bounds; (c) posterior density of β; (d) posterior density of δ; (e) joint posterior density of β and δ.

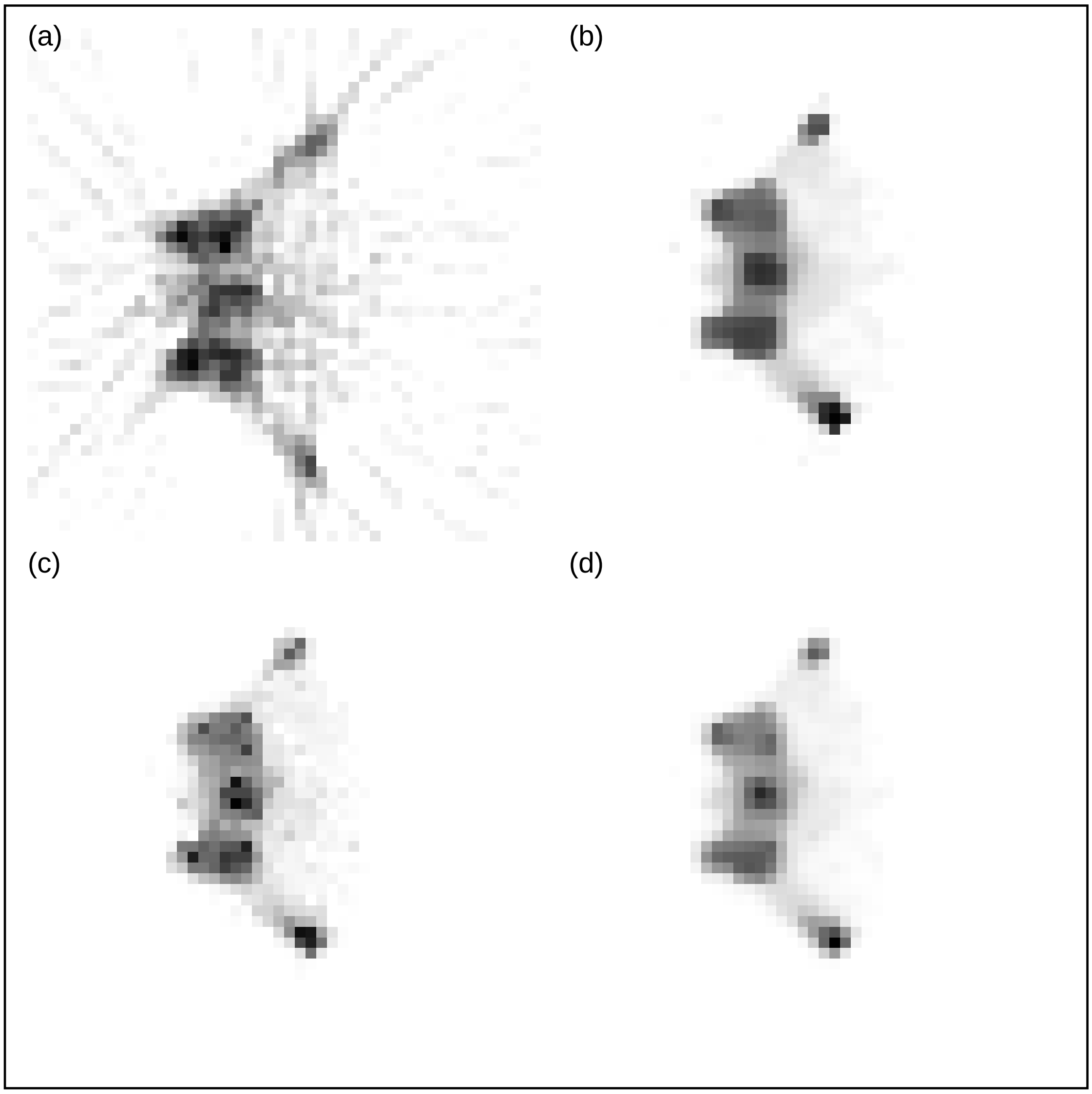

Figure 11: Reconstructions of pelvis: (a) filtered back-projection; (b) OSL posterior mode (fixed parameters); (c) realisation from fully Bayesian posterior; (d) fully Bayesian posterior mean.

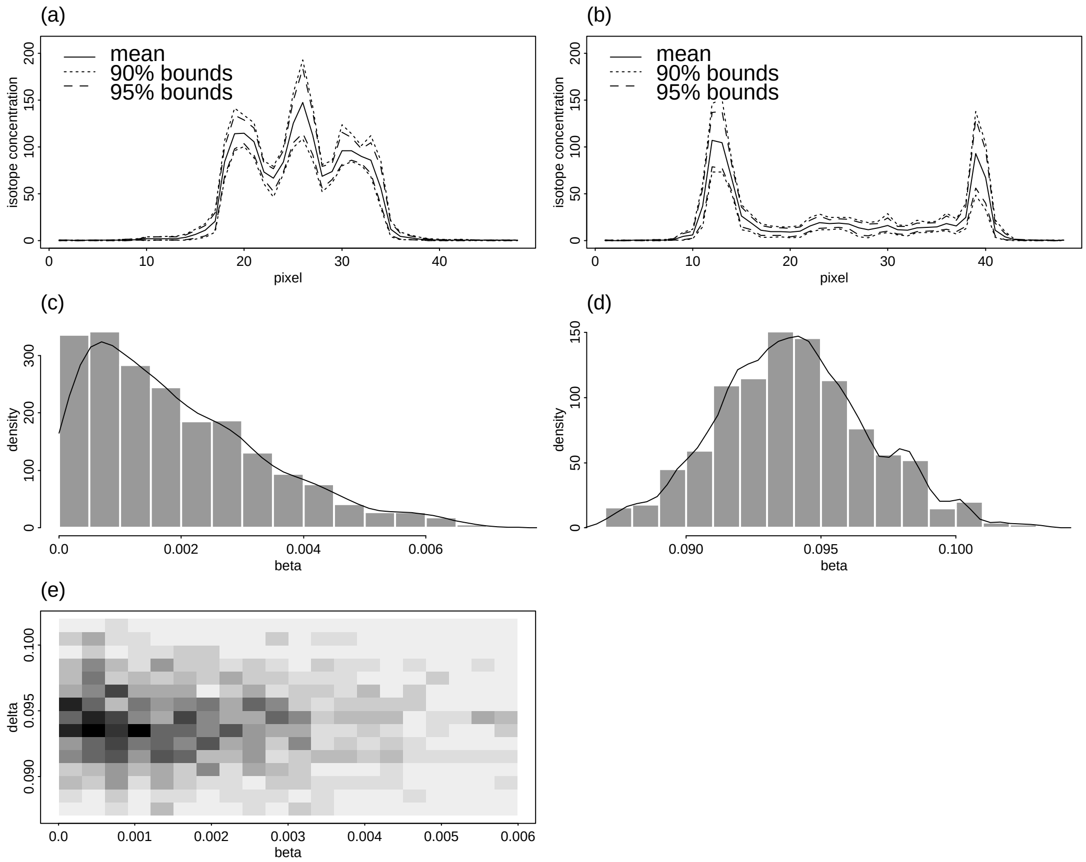

Figure 12: Pelvis: (a) column 20 posterior mean and credibility bounds; (b) column 26 posterior mean and credibility bounds; (c) posterior density of β; (d) posterior density of δ; (e) joint posterior density of β and δ.