Bayesian Model Comparison for Compartmental Models with Applications in Positron Emission Tomography

Yan Zhou, John A. D. Aston and Adam M. Johansen

Department of Statistics, University of Warwick, Coventry, CV4 7AL, UK

Summary. We develop strategies for Bayesian modelling as well as model comparison, averaging and selection for compartmental models with particular emphasis on those which occur in the analysis of Positron Emission Tomography (PET) data. Both modelling and computational issues are considered.
It is shown that an additive normal error structure does not describe measured PET data well and that within a simple Bayesian framework simultaneous parameter estimation and model comparison can be performed with a more general noise model. The proposed methodology is compared to standard techniques using both simulated and real data. In addition to good estimation performance, the proposed technique provides, automatically, a characterisation of the uncertainty in the resulting estimates which can be considerable in applications such as PET.

Keywords: Model Selection; Model Averaging; Compartmental Models

1. Introduction

In a very wide range of scientific situations, the comparison of different candidate models for observed data to assess the relative compatibility of models for data, to permit Bayesian model averaging or to perform model selection, is necessary. Various factors can make the model comparison process difficult: the scarcity of data and the presence of unknown parameters are two common difficulties and both are relevant in the context of compartmental models. For example, when analysing Positron Emission Tomography (PET) data acquired from the brain, a topic of substantial interest to neuroscientists, the number of observations available in each time course is usually between twenty and thirty, while there can be ten or more parameters to be estimated (for example, see Mankoff et al. (1998)).

The current work studies the application of Bayesian statistical methods to parameter estimation, model comparison, and model selection for compartmental models. Those compartmental models which arise in PET applications are of particular interest and are studied in greater depth in the latter part of the paper. The combination of maximum likelihood parameter estimation and either the Akaike Information Criterion (AIC), the Bayesian Information Criterion (BIC) or one of their variants for model selection is the most common approach in this field (Turkheimer et al., 2003); we compare these strategies with the proposed approach using both simulated data and real data from a PET $[^{11}\text{C}]$diprenorphine study.

Although compartmental models arise also in numerous other areas and have been extensively studied, it would seem that there are substantial differences in the inferential questions of interest. The inference of interest in the context of PET is introduced in the next section together with other general background material.

2. Background

2.1. Compartmental models

Compartmental models are a class of models which describe systems in which some real or abstract quantity flows between different (physical or conceptual) compartments, each with its own characteristics. It is often of interest to infer both parameters which describe the dynamics of the system and the number of compartments which are required in order to adequately describe measured data within this framework.

A compartmental system comprises a finite number of macroscopic subunits called compartments, each of which is assumed to contain homogeneous and well-mixed material. The compartments interact by material flowing from one compartment to another. There may be flows into one or more compartments from outside the system (inflows) and there may be flows from one or more compartments out of the system (outflows) (Jacquez, 1996). In this paper, linear compartmental models are considered, in particular those which are identifiable in PET studies (Schmidt, 1999). In these models the rate of fluid flow from a compartment is proportional to the quantity of fluid in that compartment. In such models the flow may be parameterised by a pair of transfer coefficients, which are termed rate constants and may take the value zero, for each pair of compartments.

This class of models yields a set of ordinary differential equations which describe the flow of fluid. Consider an m-compartment model. Let $\boldsymbol{f}(t)$ be the vector whose i^{th} element corresponds to the concentration in the i^{th} compartment at time t. Let $\boldsymbol{b}(t)$ describe all flow into the system from outside. The i^{th} element of $\boldsymbol{b}(t)$ is the rate of inflow into the i^{th} compartment from the environment. The dynamics of such a model may be written as:

$$\dot{\boldsymbol{f}}(t) = \boldsymbol{A}\boldsymbol{f}(t) + \boldsymbol{b}(t), \tag{1}$$

$$\boldsymbol{f}(0) = \boldsymbol{\xi}, \tag{2}$$

where $\boldsymbol{\xi}$ is the vector of initial concentrations and $\dot{\boldsymbol{f}}$ denotes the time derivative of $\boldsymbol{f}$. The matrix $\boldsymbol{A}$ is formed from the rate constants (see Gunn et al. (2001)). The solution to this equation is,

$$\boldsymbol{f}(t) = e^{\boldsymbol{A}t}\boldsymbol{\xi} + \int_0^t e^{\boldsymbol{A}(t-s)}\boldsymbol{b}(s)\,\mathrm{d}s, \tag{3}$$

where the matrix exponential $e^{\boldsymbol{A}t} = \sum_{k=0}^{\infty} \frac{(\boldsymbol{A}t)^k}{k!}$.

PET is an analytical imaging technology which uses compounds labelled with positron emitting radionuclides as molecular tracers to image and measure biochemical process *in vivo*. In a typical molecular assay, a positron-labelled tracer is injected intravenously and the PET camera scans a record of positron emission as the tracer decays (Phelps, 2000). With all events detected by the PET camera, the time course of the tissue concentrations are reconstructed as three-dimension images (Kinahan and Rogers, 1989). The digital image so captured shows the signal integrated over small volume elements (voxels). Each voxel has a volume of the order of a few cubic millimetres. This data provides the tissue time-activity function, which is the total concentration of tracer in all tissue compartments. In the *plasma input compartmental model*, in addition to the PET data, a separate measurement of the concentration of tracer in the plasma is available. This measurement is generally assumed to be noise free (it can be measured with much greater accuracy than the signal of interest). This model is used in the current study. It is assumed that the input is the same at all voxels of the reconstructed image, but each voxel can have different number of compartments associated with it. In the model fitting, a "mass univariate" approach is taken with each voxel being analysed separately. See Gunn et al. (2001) for details of PET compartmental models in general.

A plasma input model with m tissue compartments can be written as a set of ordinary differential equations,

$$\dot{\boldsymbol{C}}_{\boldsymbol{T}}(t) = \boldsymbol{A}\boldsymbol{C}_{\boldsymbol{T}}(t) + \boldsymbol{b}C_P(t) \tag{4}$$

$$C_T(t) = \mathbf{1}^T\boldsymbol{C}_{\boldsymbol{T}}(t) \tag{5}$$

$$\boldsymbol{C}_{\boldsymbol{T}}(0) = \mathbf{0}, \tag{6}$$

where $\boldsymbol{C}_{\boldsymbol{T}}(t)$ is an m-vector of time-activity functions of each tissue compartment, $C_P(t)$ is the plasma time-activity function, i.e., the input function. $\boldsymbol{A}$ is the $m \times m$ state transition matrix, $\boldsymbol{b} = (K_1, 0, \ldots, 0)^T$ is an m-vector, where K_1 is the rate constant of input from the plasma into tissue. The m-vectors $\mathbf{1}$ and $\mathbf{0}$ correspond to the vectors of ones and zeroes, respectively. The matrix $\boldsymbol{A}$ takes the form of a diagonally dominant matrix with non-positive diagonal elements and non-negative off-diagonal elements. Furthermore, $\boldsymbol{A}$ is negative semidefinite (Gunn et al., 2001). The solution to this set of ODEs is:

$$C_T(t) = C_P(t) \otimes H_{TP}(t) = \int_0^t C_P(t-s)H_{TP}(s)\,\mathrm{d}s \tag{7}$$

$$H_{TP}(t) = \sum_{i=1}^{m} \phi_i e^{-\theta_i t}, \tag{8}$$

where $\otimes$ is the convolution operator and the ϕ_i and θ_i parameters are functions of the rate constants. The input function $C_P(t)$ is assumed to be nearly continuously measured. The tissue time-activity function $C_T(t)$ is measured discretely, leading to measured values of the integral of the signal over each of n consecutive, non-overlapping time intervals ending at time points $t_1, \ldots, t_n$. The macro parameter of interest is the *volume of distribution*,

$$V_D = \int_0^\infty H_{TP}(t)\,\mathrm{d}t = \sum_{i=1}^{m} \frac{\phi_i}{\theta_i}. \tag{9}$$

This corresponds to the steady state ratio of tissue concentration to plasma concentration in a constant plasma concentration regime. That is, if an injection of tracers into the plasma were made such that the plasma concentration remained constant over the time, then the ratio of concentration in the tissues to the concentration in the plasma after an infinite time had passed would be exactly V_D.

The goal of the current work is to obtain Bayesian estimates of this macro parameter and also to estimate the posterior probabilities of models with different number of tissue compartments. Within this framework, it is possible to estimate parameters and simultaneously to deal with the number of compartments via model comparison, averaging or selection depending upon the inferential task of interest.

2.2. Bayesian model selection

When dealing with compartmental models some prior knowledge is almost always available, arising from the biophysical understanding of the system at hand. As data is generally sparse in these problems, the opportunity of making use of this information is appealing; as is the possibility of simultaneous model selection and parameter estimation.

Bayesian approaches to model selection, comparison and averaging amongst some finite collection $\mathcal{M} = \{M_1, \ldots, M_m\}$ are based upon the posterior model probability, $P(M_i|D)$, i.e. the posterior probability that model M_i is the "correct" one given that data D is observed. Simple application of Bayes rule yields $p(M_i|D) = p(D|M_i)p(M_i)/\sum_i p(D|M_i)p(M_i)$. In principle,

these probability distributions allow inference to be conducted by considering expected losses; in practice, models choice is often performed by considering the posterior mode (which is, of course, optimal under a zero-one loss function). We refer the reader to Robert (2007, chapter 7) for a discussion of these issues.

In what follows we consider a prototypical parametric model M,

$$p(D|M) = \int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} p(D|\boldsymbol{\theta}, M)p(\boldsymbol{\theta}|M)\,\mathrm{d}\boldsymbol{\theta}, \tag{10}$$

where $\boldsymbol{\theta}$ is the parameter vector and $\boldsymbol{\Theta}$ is the parameter space of model M. The model specifies the likelihood function $p(D|\boldsymbol{\theta}, M)$ and prior beliefs are expressed through the prior distribution $p(\boldsymbol{\theta}|M)$. Given a prior over the models and a prior distribution for the parameters of each model, Bayesian model comparison proceeds via the calculation of the marginal likelihoods $p(D|M)$. It is well known that the prior specified over model parameters can substantially alter the posterior model probabilities (cf. Kass and Raftery (1995)).

Calculating Marginal Likelihoods In most realistic situations $p(D|M)$ cannot be obtained analytically. However, the posterior density, $p(\boldsymbol{\theta}|D, M)$, is proportional to $p(D|\boldsymbol{\theta}, M)p(\boldsymbol{\theta}|M)$ with the normalizing constant equal to $p(D|M)$. Therefore Monte Carlo methods are widely used to provide sample approximations of the posterior distribution; the marginal likelihood can be estimated using these sample approximations.

Markov chain Monte Carlo The principle of Markov chain Monte Carlo (MCMC) is that the sequence of dependent random variables, $\{X^{(i)}\}_{i\geq 1}$, produced by an ergodic Markov chain with invariant distribution f provides a Monte Carlo approximation of the integral $\int h(x)f(x)\,\mathrm{d}x$ where $h(x)$ is any sufficiently regular function in the form of the ergodic average:

$$\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^{n} h(X^i) \to \mathbb{E}_f[h(X)]$$

see (Tierney, 1994; Robert and Casella, 2004).

Suppose an MCMC algorithm with invariant distribution $p(\boldsymbol{\theta}|D) \propto p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})$ is available; it will produce a sequence of dependent samples for parameter $\boldsymbol{\theta}$, $(\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(T)})$ From the identity

$$\int_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} g(\boldsymbol{\theta})\frac{p(\boldsymbol{\theta}|D)p(D)}{p(D|\boldsymbol{\theta})p(\boldsymbol{\theta})}\,\mathrm{d}\boldsymbol{\theta} = 1, \tag{11}$$

where g is any probability density function whose support is contained within that of the posterior. It follows that, under weak regularity conditions, a consistent estimator of $p(D)$ is

$$\widehat{p(D)} = \left[\frac{1}{m}\sum_{i=1}^{m} \frac{g(\boldsymbol{\theta}^{(i)})}{p(D|\boldsymbol{\theta}^{(i)})p(\boldsymbol{\theta}^{(i)})}\right]^{-1}. \tag{12}$$

This estimator obeys a central limit theorem if the tails of g are sufficiently light. To avoid instability arising from samples with very small likelihood, g should be chosen to have lighter tails than the posterior distribution (Congdon, 2006). Note that these requirements are exactly those that arise in importance sampling (cf. Geweke (1989)) although in this setting we have freedom to specify the target rather than the proposal density.

2.3. Information Criteria

AIC and BIC are information criteria which are widely used for model selection when point estimates of parameters are available; their use is ubiquitous in the analysis of PET data. Both rely on the asymptotic behaviour of maximum likelihood estimator (MLE).

The AIC was introduced by Akaike (1973). The principle idea is to select the model which minimises amongst the class of candidate models, asymptotically, the Kullback-Leibler divergence (Kullback and Leibler, 1951), between the true density of the data and that of the model. In this approach, the preferred model is that which minimises $\text{AIC} = -2\widehat{\ell} + 2k$, with $\widehat{\ell}$ denoting the maximum of the log likelihood and k the number of estimated parameters in the model. A small sample correction, $\text{AIC}' = -2\widehat{\ell} + 2k + 2k(k-1)/(n-k-1)$ suitable for samples of size $n \lessapprox 50k$ was proposed by Hurvich and Tsai (1989). This is the expansion which was used in the analysis below.

The BIC was developed by Schwarz (1978) based upon a large sample approximation of the Bayes factor. Defined as $\text{BIC} = -2\widehat{\ell} + k\ln(n)$, an asymptotic argument concerning Bayes factors under appropriate regularity conditions justifies the choice of the model with the smallest value of BIC.

2.4. Connections with Existing Work

A great deal of work has been done on the analysis of compartmental models and also of PET data; the current section summarises the relationship between the current work and the most relevant parts of this literature.

The use of AIC based methods for compartmental models was first introduced by Hawkins et al. (1986). Their work, and some recent use of AIC focus on low noise data (for example Turkheimer et al. (2003) used AIC for model averaging for region of interest data). In our case, the data has much higher noise and it seems to violate the conditions which are required to justify the use of AIC to a considerable extent. The model has a nonlinear structure and the high level noise cannot easily be approximated by normal distribution. In such situations, it was shown that indeed for either simulated data or real data, AIC does not perform well. Thus, Bayesian modelling is the focus in this work.

We note that Bayesian analysis of compartmental models has been considered extensively in other application domains with considerable success. In particular, the Bayesian analysis of compartmental models in pharmacokinetics has received considerable attention since the work of Wakefield (1996) and much work has been done on the analysis of related models in epidemiology (e.g. Gibson and Renshaw (1998)). However, in these areas the questions of interest have typically been different; when model selection has been considered in the case of pharmacokinetics the object of inference has typically been considering which covariates to include in a regression analysis whilst in the epidemiological setting, variable dimension models arise from considering interactions between individuals and subpopulations. In both cases the number of compartments is typically treated as known and ascribed a particular physical significance. This is quite different from the setting considered here.

3. Methodology

3.1. Models

In the scenarios below, linear one-, two-, and three-compartment models are considered possible; the method could deal with other compartmental models straightforwardly but we focus on these as they are the most interesting in the application of interest. Let $t_1, \ldots, t_n$ be the end point of time frames at which the tissue concentrations are measured, let y_j, $j = 1, \ldots, n$ be the observed

data. Measurement error is assumed to be white and additive with zero mean and variance proportional to activities divided by the length of time frames (these assumptions arise from the physical characterisation of the PET system of interest; alternative specifications would be possible and appropriate for other situations). Recall equation (7),

$$C_T(t_j;\phi_1,\theta_1,\ldots,\phi_m,\theta_m) = \sum_{i=1}^{m}\phi_i\int_0^{t_j} C_P(s)e^{-\theta_i(t_j-s)}\,\mathrm{d}s$$

$$y_j = C_T(t_j;\phi_1,\theta_1,\ldots,\phi_m,\theta_m) + \sqrt{\frac{C_T(t_j;\phi_1,\theta_1,\ldots,\phi_m,\theta_m)}{t_j-t_{j-1}}}\varepsilon_j,$$

where $m = 1, 2,$ or 3 is the number of tissue compartments, $t_0 = 0$, and ε_j's are identically independently distributed random variables with mean zero. It is usually assumed that ε_j's have a normal distribution. It is demonstrated below that there is evidence that a t distribution better fits the observed data. We consider both error structures:

$$\varepsilon_j \sim \mathcal{N}(0,\sigma^2) \qquad \text{Normal distributed errors} \tag{13}$$

$$\varepsilon_j \sim \mathcal{T}(0,\tau,\nu) \qquad \text{t distributed errors,} \tag{14}$$

where $\mathcal{N}(0,\sigma^2)$ is the normal distribution with mean zero and variance σ^2, and $\mathcal{T}(0,\tau,\nu)$ is the Student t distribution with location zero, scale τ, and degrees of freedom ν.

3.2. NLS, AIC, and BIC implementations

The AIC and BIC approaches to model comparison are both based upon the maximum likelihood estimates (MLE). With normally distributed errors, the log likelihood with respect to the ϕ_j, θ_j's and σ^2 parameters is,

$$\begin{aligned}\ell =& \frac{n}{2}\ln\Big(\frac{1}{2\pi\sigma^2}\Big) + \frac{1}{2}\sum_{j=1}^{n}\Big(\frac{t_j-t_{j-1}}{C_T(t_j;\phi_1,\theta_1,\ldots,\phi_m,\theta_m)}\Big)\\ &- \frac{1}{2\sigma^2}\sum_{j=1}^{n}\frac{t_j-t_{j-1}}{C_T(t_j;\phi_1,\theta_1,\ldots,\phi_m,\theta_m)}(y_j - C_T(t_j;\phi_1,\theta_1,\ldots,\phi_m,\theta_m))^2,\end{aligned} \tag{15}$$

and given values for the ϕ and θ parameters, ℓ is maximised by

$$\widehat{\sigma^2} = \sum_{j=1}^{n}\frac{t_j-t_{j-1}}{C_T(t_j;\hat\phi_1,\hat\theta_1,\ldots,\hat\phi_m,\hat\theta_m)}(y_j - C_T(t_j;\hat\phi_1,\hat\theta_1,\ldots,\hat\phi_m,\hat\theta_m))^2, \tag{16}$$

where $C_T(t_j)$ is evaluated at the estimates of ϕs and θs. The nonlinear least squares (NLS) method for approximation of the MLE is widely used in PET models; particularly in the neuroscience literature. Throughout the current work, NLS, AIC, and BIC are implemented such that, first the estimates of ϕs and θs are found by minimising

$$\sum_{j=1}^{n}\frac{t_j-t_{j-1}}{C_T(t_j;\hat\phi_1,\hat\theta_1,\ldots,\hat\phi_m,\hat\theta_m)}(y_j - C_T(t_j;\hat\phi_1,\hat\theta_1,\ldots,\hat\phi_m,\hat\theta_m))^2. \tag{17}$$

Then the maximised log likelihood, which is required by AIC and BIC, is evaluated with the NLS estimates for ϕs and θs, together with the estimator for σ^2 as in equation (16). This approximation of the MLE is widely used in the literature (Turkheimer et al., 2003) because it's

rather easy to compute, but it can exhibit somewhat unstable behaviour at high noise levels. See, for example, the simulation study of Peng et al. (2008).

When implementing the NLS algorithm, the ϕ parameters are constrained to lie within the interval $[10^{-5}, 1]$ and the θ parameters within the interval $[10^{-4}, 1]$ in order to ensure that the parameters are physiologically meaningful.

3.3. Bayesian modelling for PET Compartmental Models

We consider Bayesian models for the observed data signal under the hypotheses that residual noise is well modelled by (a) additive normal errors and (b) additive t distributed errors.

It is necessary to specify prior distributions for the ϕ and θ parameters. Two approaches are employed; this enables the assessment of the sensitivity of our results to prior specification. We consider vague priors in which scale parameters follow an approximation to the Jeffrey's prior and rate constants are assumed to follow a uniform distribution on the same intervals as are considered feasible in the NLS implementation. We also consider prior distributions informed by biological knowledge as discussed in the next section.

With normally distributed errors, the prior for precision parameter $\lambda = \frac{1}{\sigma^2}$ is a gamma distribution with both parameters equal 10^{-3} – a proper approximation to the improper Jeffrey's prior for scale-parameter λ, $p(\lambda) \propto \frac{1}{\lambda}$. With t distributed errors, the same prior is used for the scale parameter, τ, as for λ in the normal model. The prior for $1/\nu$ is uniform over interval $[0, 0.5)$, allowing the likelihood to vary from having a very heavy tail to being arbitrarily close to normality (Gelman et al., 2004). Let $\boldsymbol{y} = (y_1, \ldots, y_n)^T$, and recall that

$$C_T(t_j; \phi_1, \theta_1, \ldots, \phi_m, \theta_m) = \sum_{i=1}^{m} \phi_i \int_0^{t_j} C_P(s) e^{-\theta_i (t_j - s)} \, \mathrm{d}s.$$

Using the above priors, the posterior distribution with normally distributed errors is

$$\begin{aligned} p(\phi_1, \theta_1, \ldots, \phi_m, \theta_m, \lambda | \boldsymbol{y}) \propto & \prod_{j=1}^{n} \sqrt{\lambda} \exp\Big\{ -\frac{\lambda}{2} (y_j - C_T(t_j; \phi_1, \theta_1, \ldots, \phi_m, \theta_m))^2 \Big\} \\ & \times \lambda^{\alpha - 1} e^{-\beta \lambda} \prod_{i=1}^{m} I_{[\phi_i^a, \phi_i^b]}(\phi_i) I_{[\theta_i^a, \theta_i^b]}(\theta_i), \end{aligned} \tag{18}$$

where $\alpha = \beta = 10^{-3}$, the parameters of the prior distribution of λ. And ϕ_i^a and ϕ_i^b are the lower and upper bounds of the truncation interval of parameter ϕ_i and corresponding notation is used for θ_i. These intervals are the same as those used to constrain the NLS estimates for these parameters.

With t distributed errors, y_j has a t distribution with location $C_T(t_j)$, scale $\frac{t_j - t_{j-1}}{C_T(t_j)} \tau$, and degrees of freedom ν. The posterior distribution is,

$$\begin{aligned} & p(\phi_1, \theta_1, \ldots, \phi_m, \theta_m, \tau, \nu | \boldsymbol{y}) \\ & \propto \prod_{j=1}^{n} \Bigg\{ \frac{\Gamma(\frac{\nu+1}{2})}{\Gamma(\frac{\nu}{2})} \Big(\frac{t_j - t_{j-1}}{C_T(t_j; \phi_1, \theta_1, \ldots, \phi_m, \theta_m)} \frac{\tau}{\pi \nu} \Big)^{\frac{1}{2}} \\ & \times \Big(1 + \frac{t_j - t_{j-1}}{C_T(t_j; \phi_1, \theta_1, \ldots, \phi_m, \theta_m)} \frac{\tau}{\nu} (y_j - C_T(t_j; \phi_1, \theta_1, \ldots, \phi_m, \theta_m))^2 \Big)^{-\frac{\nu+1}{2}} \Bigg\} \\ & \times \tau^{\alpha - 1} e^{-\beta \tau} \times \frac{1}{\nu^2} \times I_{[a,b]}\Big(\frac{1}{\nu}\Big) \prod_{i=1}^{m} I_{[\phi_i^a, \phi_i^b]}(\phi_i) I_{[\theta_i^a, \theta_i^b]}(\theta_i) \end{aligned} \tag{19}$$

where $\alpha = \beta = 10^{-3}$, the parameters of the prior distribution of τ; $a = 0$ and $b = 0.5$.

3.3.1. Biologically informed priors

The primary prior information when dealing with compartmental models typically concerns the macro parameter(s) of interest: V_D in the situations considered here. However, it is more convenient to work with models expressed in terms of the collection $\{\theta_i, \phi_i\}_{i=1}^m$. Here, a method for constructing informative priors in terms of these parameters is presented.

Anderson (1983) provided some useful results about compartmental models in general. Let γ_{0j} denote the rate constant of the outflow from the j^{th} compartment into the environment. Without loss of generality, assume that the θ_i are ordered: $\theta_1 \leq \cdots \leq \theta_m$. Then,

(a) $0 \leq \theta_i \leq 2 \max_j |A_{jj}|$ for all i.
(b) $\min_j \gamma_{0j} \leq \theta_1 \leq \max_j \gamma_{0j}$.
(c) when there is only one outflow into the environment, say the rate constant of this outflow is k_2, as in the plasma input model, then $0 \leq \theta_1 \leq k_2$.

In addition, $\sum_{i=1}^m \phi_i = K_1$, where K_1 is the rate constant of input from the plasma into the tissues (Gunn et al., 2001). Therefore $\phi_i < K_1$ for $i = 1, \ldots, m$. Given this information, more informative prior distributions can be constructed. For simplicity, we restrict discussion to imposing upper and lower bounds on the possible values of the parameters.

To demonstrate the idea, an informative prior distributions for a three tissue compartments model is constructed. First note that the transition matrix A is,

$$\boldsymbol{A} = \begin{bmatrix} -k_2 - k_3 - k_5 & k_4 & k_6 \\ k_3 & -k_4 & 0 \\ k_5 & 0 & -k_6 \end{bmatrix} \tag{20}$$

It is believed that all the rate constants take values in the range $[5 \times 10^{-4}, 10^{-2}]$. Without loss of generality, we impose the identifiability constraint $\theta_1 \leq \theta_2 \leq \theta_3$, then,

$$0 < \theta_1 \leq k_2 \leq 10^{-2} \tag{21}$$

$$\theta_1 \leq \theta_2 \leq \theta_3 \leq \max\{2(k_2 + k_3 + k_5), 2k_4, 2k_6\} \leq 6 \times 10^{-2} \tag{22}$$

Under the imposed ordering, as θ_1 is the smallest exponent, the term $\phi_1 e^{-\theta_1 t}$ decays more slowly than any other term in the expansion. Consequently, ϕ_1/θ_1 is likely to make a relatively large contribution to $V_D = \sum_{i=1}^m \phi_i/\theta_i$. In fact, since A has only negative real eigenvalues, θ_1 is the spectral radius of A. But it is not well known how large the ratio $(\phi_1/\theta_1)/V_D$ will be. However it is easy to conduct a numerical study here, given the small number of parameters. It is found that among all possible combination of ks, for all the combinations, $\phi_1/\theta_1 \geq 0.5V_D$. If the combinations of ks are restricted to those without too large differences among them, that is cases like $k_5 \gg k_6$ are not considered, then $\phi_1/\theta_1 \geq 0.7V_D$. The reason for not considering these cases is that such irreversible (trapped) models yield infinite V_D estimates and tracers are generally known a-priori as to whether they are reversible or irreversible.

In summary, given the belief that the rate constants lie within $[5 \times 10^{-4}, 10^{-2}]$, the macro parameters $K_1 \sim 5 \times 10^{-3}$ml s^{-1} cm^{-3}, and $V_D \sim 20$, we will have the following beliefs:

(a) $\phi_1 \sim 0.8 \times (5 \times 10^{-3}) = 4 \times 10^{-3}$.
(b) $\phi_1/\theta_1 \sim 0.8 \times 20 = 16$.
(c) $\phi_i < K_1 - \phi_1$ for all $i > 1$.
(d) $\phi_i/\theta_i < V_D - \phi_1/\theta_1$ for all $i > 1$.

Defining the truncated normal density, using indicator function $\mathbb{I}$ and standard normal distribution function Φ:

$$\mathcal{TN}_{[a,b]}\left(x;\mu,\sigma^2\right) := \frac{\mathcal{N}(x;\mu,\sigma^2)\mathbb{I}_{[a,b]}(x)}{\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)}, \tag{23}$$

the following prior distributions are used to encode this information:

$$\phi_1 \sim \mathcal{TN}_{[10^{-5},10^{-2}]}\left(\cdot;3\times10^{-3},10^{-3}\right) \tag{24}$$
$$\theta_1|\phi_1 \sim \mathcal{TN}_{[2\times10^{-4},10^{-2}]}\left(\cdot;\phi_1/15,10^{-2}\right) \tag{25}$$
$$\phi_2 \sim \mathcal{TN}_{[10^{-5},10^{-2}]}\left(\cdot;10^{-3},10^{-3}\right) \tag{26}$$
$$\theta_2|\phi_2,\theta_1 \sim \mathcal{TN}_{[\theta_1,6\times10^{-2}]}\left(\cdot;\phi_2/4,10^{-2}\right) \tag{27}$$
$$\phi_3 \sim \mathcal{TN}_{[10^{-5},10^{-2}]}\left(\cdot;10^{-3},10^{-3}\right) \tag{28}$$
$$\theta_1|\phi_3,\theta_2 \sim \mathcal{TN}_{[\theta_2,6\times10^{-2}]}\left(\cdot;\phi_3/1,10^{-2}\right) \tag{29}$$

For one- and two-compartments models the appropriate subset of these prior distributions are used, ensuring that common priors are used for the shared parameters of nested models.

3.3.2. MCMC algorithms

The MCMC algorithm used to sample the posterior distribution is an Metropolis random walk algorithm. Let p denote the number of parameters. Let $\boldsymbol{\psi} = (\psi_1,\ldots,\psi_p)$ be the parameter vectors, which will be $(\phi_1,\theta_1,\ldots,\phi_m,\theta_m,\lambda)$ for the normal distributed errors or $(\phi_1,\theta_1,\ldots,\phi_m,\theta_m,\tau,\nu)$ for the t distributed errors. Let $f(\boldsymbol{\psi})$ be the target distribution, which is as in Equations (18) and (19) for normal and t distributed errors, respectively, when vague priors are used and the corresponding distributions with informative priors. The algorithm specifies as,

(a) Initialize $\boldsymbol{\psi}$ with $\boldsymbol{\psi}^{(0)} = \boldsymbol{\psi}_0$, set $t = 0$. $\boldsymbol{\psi}_0$ can be any value within the boundaries of the priors.
(b) Generate $\boldsymbol{U}_t$ according to p-dimensional uniform random distribution on $\prod_{i=1}^{p}[-s_i,s_i]$. Where s_i is the step size for ψ_i, which will be specified later. Set $\boldsymbol{\eta}_t = \boldsymbol{\psi}^{(t)} + \boldsymbol{U}_t$.
(c) Calculate $r_t = f(\boldsymbol{\eta}_t)/f(\boldsymbol{\psi}^{(t)})$. Generate u_t according to uniform distribution on $[0,1]$. If $u_t \le r_t$, Set $\boldsymbol{\psi}^{(t+1)} = \boldsymbol{\eta}_t$, otherwise set $\boldsymbol{\psi}^{(t+1)} = \boldsymbol{\psi}^{(t)}$. Increment t. If $t < N$ for some preset positive integer N, go to step (b), otherwise stop.

The step sizes are chosen such that the acceptance rate is controlled around 20% to 30%. The marginal likelihood is estimated with equation (12) where g is chosen to be the multivariate normal distribution, whose mean and covariance matrix are calculated using the posterior samples.

4. Numerical Results

We begin with a simulation study to validate the proposed method before moving on to consider real data from two [^{11}C]diprenorphine experiments.

4.1. One-dimension simulation

The data is simulated from the three-compartment model, with parameters $K_1 = 6\times10^{-3}$, $k_2 = 3\times10^{-3}$, $k_3 = 5.5\times10^{-3}$, $k_4 = 1.5\times10^{-3}$, $k_5 = 10^{-3}$ and $k_6 = 3\times10^{-3}$. All parameters have the unit s^{-1} except K_1 which has units ml s^{-1} cm^{-3} (Innis et al., 2007). The macro

parameter V_D is thus 10. A real measured plasma input function, taken from Jiang et al. (2009), is used (see figure 1). The simulated data has 32 time frames with lengths corresponding to the integration periods used in real experiments (27.5, 32.5, 2×10, 20, 6×30, 75, 11×120, 210, 5×300, 450, and 2×600, all in seconds), see figure 2 for the synthetic noise free data. The noise is added to the synthetic data such that the noise is normal distributed with mean zero, and variance proportional to the time activities divided by the length of time frames. The largest data point has variance scaled to unity before all of the variances are scaled in proportion to a "noise level" variable. This noise level ranges from 0.01 to 5.12, from lower than typical region of interest (ROI) analysis (in which the data is averaged over a biologically meaningful region in order to improve signal to noise ratio) to higher than the noise associated with voxel level analysis (Peng et al., 2008). For each noise level, 2,000 time series were simulated. Normally distributed errors were assumed (correctly, in this simulation study).

The NLS procedures use the DIRECT algorithm (Jones et al., 1993) to find a local minimum, and then uses this minimum to initialise a Nelder-Mead simplex algorithm (Nelder and Mead, 1965). The programs used for both likelihood-based and Bayesian modelling are implemented in C++ (Stroustrup, 1991) and are available from the first author on request.

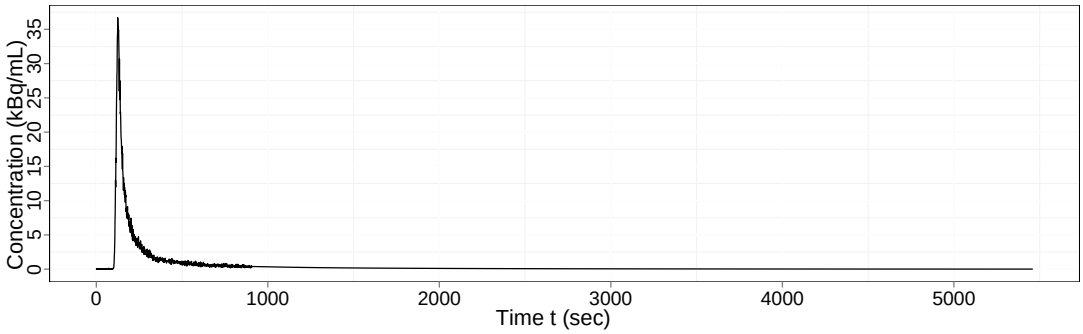

Fig. 1. Input function C_P

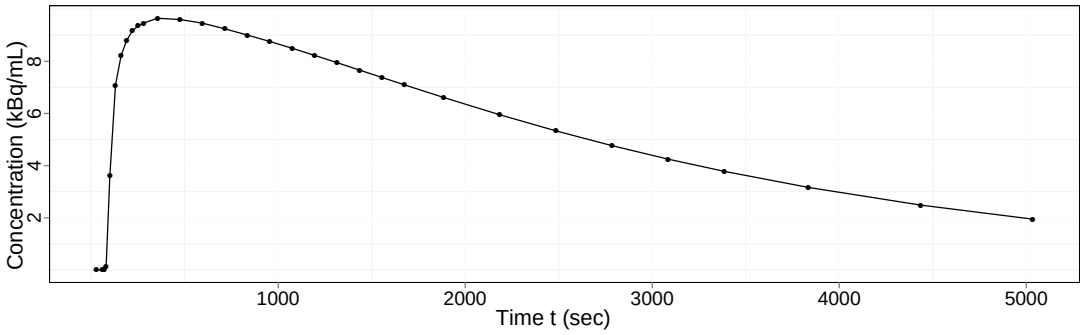

Fig. 2. Noise free simulated data

Parameter Estimates Table 1 summarises the MSE of estimates for three-compartments model, by NLS, Bayesian estimates with vague priors, and Bayesian estimates with informative priors.

Table 1. MSE of V_D, three-compartment model

	Noise level									
Method	0.01	0.02	0.04	0.08	0.16	0.32	0.64	1.28	2.56	5.12
NLS	0.0005	0.001	0.004	0.017	0.032	0.052	0.103	0.221	0.572	1.191
Bayesian vague	0.0005	0.0009	0.002	0.004	0.008	0.015	0.031	0.053	0.105	0.207
Bayesian informative	0.0004	0.0008	0.002	0.003	0.007	0.013	0.027	0.052	0.104	0.195

Table 2. Frequencies of model selected by AIC and BIC (%)

		Noise level									
	Model	0.01	0.02	0.04	0.08	0.16	0.32	0.64	1.28	2.56	5.12
AIC	1	0	0.1	0.6	1.0	1.8	16.3	48.8	78.3	91.6	98.5
	2	91.6	94.0	95.0	96.3	96.6	83.1	50.7	21.5	8.3	2.5
	3	8.4	5.9	4.4	2.7	1.6	0.6	0.5	0.2	0.1	0
BIC	1	0	0.1	0.8	1.3	3.5	27.1	64.9	87.8	95.7	98.6
	2	94.6	96.2	96.1	96.8	95.5	72.7	35.0	12.2	4.3	1.4
	3	5.4	3.7	3.1	1.9	1.0	0.2	0.1	0	0	0

As shown in the table, the NLS estimates are good at low noise level. But at the noise levels typically observed in voxel-level analyses, the Bayesian estimates have significantly smaller MSE. The estimates with informative priors improves at low noise level and are comparable to the estimates with uniform priors at high noise level. In general the Bayesian estimates are more stable than the NLS estimates, which is known to have positive bias which increases with noise level (Peng et al., 2008).

Model selection Tables 2 and 3 summarise the proportion of times each order of model is selected by the information criteria techniques and by choosing the *a posteriori* most probable model with a uniform prior over model order, respectively. Table 4 summarises the MSE of estimates using selected model under these model selection strategy. As shown in the table, both the frequency with which the true model is chosen and the MSE of estimated for selected models are improved by using Bayesian model selection. Model selection is improved, particularly at higher noise levels, by the use of informative priors. However, in all cases, the true model is hard to identify due to the limited temporal data, even at low noise levels.

Table 3. Frequencies of model selected by Bayes factors (with uniform and informed priors) (%)

		Noise level									
	Model	0.01	0.02	0.04	0.08	0.16	0.32	0.64	1.28	2.56	5.12
Vague	1	0	0	0	0	6.3	7.0	24.3	30.7	41.6	54.8
Priors	2	12.5	20.1	35.2	49.4	55.3	67.5	62.6	59.1	52.2	43.0
	3	87.5	79.9	64.8	50.6	38.4	25.5	13.1	10.2	6.2	2.2
Informative	1	0	0	0	0	0	1.0	6.2	15.2	27.8	37.1
Priors	2	10.6	17.5	33.3	45.8	58.8	70.2	73.0	67.3	57.3	53.0
	3	89.4	82.5	66.7	54.2	41.2	28.8	20.8	17.5	14.9	9.9

Table 4. MSE of V_D, selected model

	Noise level									
Method	0.01	0.02	0.04	0.08	0.16	0.32	0.64	1.28	2.56	5.12
AIC	0.0005	0.001	0.003	0.007	0.012	0.024	0.063	0.132	0.308	0.719
BIC	0.0005	0.001	0.002	0.006	0.011	0.018	0.059	0.111	0.242	0.658
BF Vague	0.006	0.006	0.007	0.009	0.022	0.025	0.031	0.074	0.085	0.247
BF Informative	0.001	0.001	0.002	0.004	0.007	0.015	0.028	0.058	0.111	0.221

Discussion Using different priors does not alter the results substantially: there is sufficient information in the data to overwhelm the quite large differences between the two sets of priors which were considered. Model selection is not overly sensitive to the choice of priors (although slightly better results are obtained by using informative priors). Bayesian modelling is overall better than using AIC or BIC combined with NLS, for both parameter estimation and model selection in that it produces estimates with a smaller MSE for most noise levels and recovers the true model more often than the alternative approaches. For model selection with very noisy data, the AIC and BIC methods can hardly detect the existence of second compartment and at no noise level can they find considerable evidence of the existence of the third compartment (which exists in the true model). The Bayesian approach to model selection shows a large improvement over AIC and BIC, but still cannot recover the true model reliably. It is our view that this provides a strong motivation for treating model selection very cautiously when dealing with models of this sort and that the Bayesian approach developed has two substantial advantages: it provides a natural quantification of uncertainty and, perhaps more significantly, it lends itself to a model averaging approach which is perhaps more appropriate when no one model is overwhelmingly preferred.

4.2. Measured [^{11}C]diprenorphine data

Having verified that the proposed method is effective when applied to data simulated from the model, we turn our attention to real data sets which have been considered in the literature using this model.

Data from the a PET study using [^{11}C]diprenorphine are used to examine the methods presented. The overall aim of the study was to quantify opioid receptor concentration in the brain of normal subjects allowing a baseline to be found for subsequent studies on diseases such as epilepsy. The data have been previously analysed in Peng et al. (2008) and in Jiang et al. (2009) but in both these previous analyses, parameter estimation was the focus rather than model comparison. Two dynamic scans from a measured [^{11}C]diprenorphine study of normal subjects, for which an arterial input function was available, were analysed. [^{11}C]diprenorphine is a tracer which binds to the opioid (pain) receptor system in the brain. The subjects underwent 95-min dynamic [^{11}C]diprenorphine PET baseline scans on the same camera. The subjects were injected 185 MBq of [^{11}C]diprenorphine. PET scans were acquired in 3D mode on a Siemens/CTI ECAT EXACT3D PET camera, with a spatial resolution after image reconstruction of approximately 5mm. Data were reconstructed using the reprojection algorithm (Kinahan and Rogers, 1989) with ramp and Colsher filters cutoff at Nyquist frequency. Reconstructed voxel size were 2.096mm $\times$ 2.096mm $\times$ 2.43mm. Acquisition was performed in listmode (event-by-event) and scans were rebinned into 32 time frames of increasing duration. Frame-by-frame movement correction was performed on the PET images.

Nonnegative least squares (NNLS) estimates of V_D (Cunningham and Jones, 1993) are available from a previous study for both data sets and are used as a baseline for comparison (NNLS can

be used due to the non-negative nature of the underlying rate constants). The AIC and BIC strategies select the model with smallest AIC or BIC, respectively while the Bayesian strategy selects model with highest marginal likelihood. The NLS procedures are exactly the same as in the simulation study. One-, two- and three- tissue compartment models are fitted. The same models (with normally distributed errors) were also subjected to Bayesian analysis. However, the results are not reasonable. Figure 3 shows the time series associated with five typical voxels and their standardised residuals (which in the case of a normal error model should have a standard normal distribution). The NNLS estimates for these five voxels are all around 25 or above. But the Bayesian estimates with normally distributed errors are about 15. In fact, for most voxels, the Bayesian estimates are about 50% smaller than the NNLS estimates. This can be explained by noting that for the first three data observations, the input function is nearly zero. Hence whatever values of parameters are proposed, the fitted value of $C_T(t)$ will be near zero for these three points (no input, no activity). Whenever any of the first three observation departs significantly from zero, which for the noisy data at voxel level analysis happens in almost all real data, the likelihood will be very low. The normal distribution has a very thin tail and therefore the first few data points have an overwhelming effect on the likelihood and hence the posterior distribution. In practical studies, the influence of these points is truncated at a somewhat arbitrary value (a maximum weight of 1,000 was used within the NLS component of our own comparison) in order to mitigate against this effect: this ad hoc procedure is essential in order to obtain reasonable results with this model.

As shown in the figure, all of the five typical voxels have large residuals at the start of the time-activity course. With normal distributed errors, these points will have very small probability. Testing the residuals against normal distribution with Kolmogorov-Smirnov test (which is overly conservative as the parameters of the normal distribution have been estimated) shows that for the great majority of the voxels across the whole space, the null hypothesis (that residuals are from a normal distribution) should be rejected at a 5% level. A possible solution to this problem is proposed here. The t distribution is used in place of the normal distribution to model the errors. The t distribution can have a heavy tail and is more robust to outliers than the normal distribution. As shown in figure 3, the data residuals demonstrate rather systematic and substantial departures from the normal distributions.

Diagnostics for the Convergence of the Markov Chain For all data sets, 100,000 iterations are used for burn-in and a further 200,000 iterations are used to make the subsequent inference. The estimates of V_D and the marginal likelihood $p(D)$ are our primary objects of inference. In order to assess the convergence, the MCMC chain was initialised with dispersed starting values. Figure 4 shows the estimates of V_D from the burn-in iterations of a typical voxel when starting the chain from different values. As shown in the plot, 40,000 iterations is enough for the chain to mix well and get a good estimate for V_D, the parameter of interest, as mentioned above, 100,000 samples were used for a conservative burn-in period. Similar plots are produced for other parameters. And all show that initial the chains from different areas of the parameter space does not change the estimates. It is not feasible manually inspect such traces for all voxels, instead, 200 voxels with a range of values of V_Ds were examined. It was found that the algorithm mixed well for voxels from different regions of the brain.

In addition, much longer chains were run to examine the behaviour of the algorithm. Table 5 shows the estimates of a typical voxel when using different length of the MCMC chain. As seen in the table, with long chains, the estimates does not change substantially suggesting that the algorithm has converged to stationarity for even the shortest chains (of course, no such diagnostic provides *proof* that this is the case).

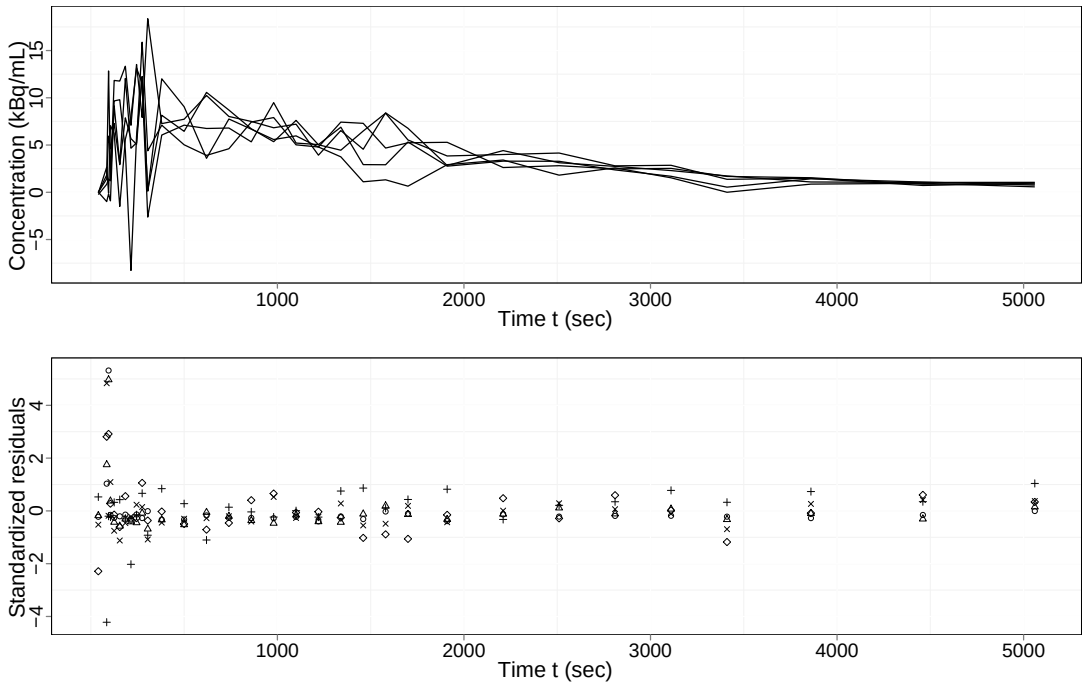

Fig. 3. Measured $C_T(t)$ for five typical voxels and their standardized residuals, fitted with NLS for three-compartments models

Table 5. Estimates from a long chain for a typical voxel

	Chain lengths					
Parameter	10^4	5×10^4	10^5	5×10^5	10^6	2×10^6
V_D	25.71	25.67	25.64	25.75	25.82	25.73
$p(D)/10^{-31}$	3.22	3.45	3.36	3.23	3.31	3.39

Estimates Figure 5 shows the estimates of V_D for a three-compartments model, using Bayesian posterior means with informative priors and NLS, together with the NNLS estimates obtained by Jiang et al. (2009) for the data. Overall the percentage difference between NNLS and NLS ((NNLS - NLS) / NLS) is about 3%. This difference is fairly uniform overall the range of V_D, though there are large percentage difference for voxels with very small values of V_D. This is due to the fact that different bounds/priors are applied to the space of the ϕ and θ parameters and different estimates are obtained when the value of parameters are near the boundaries. However, these voxels are less of interest as they correspond to regions with little or no activity and hence little receptor density. The Bayesian estimates are roughly 5% smaller than the NNLS estimates. Previous studies showed that NNLS has about 5% positive bias with noise levels typical of a voxel analysis (Peng et al., 2008). There are similar large differences for voxels with small V_D as with the NLS estimates. Overall, if we take the results of previous simulation study of NNLS, then the Bayesian estimates would appear to offer better estimation than NNLS estimates. In addition, a principled way of introducing the prior knowledge concerning the rate constants has been used and a more appropriate noise model has been employed aiding interpretation of the

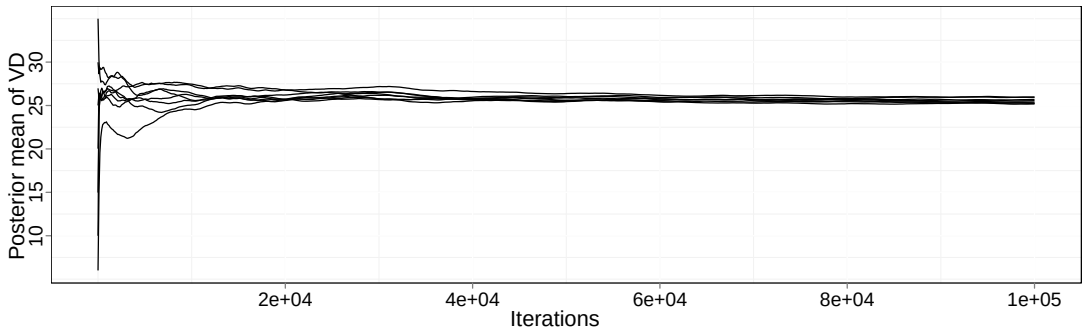

Fig. 4. Estimates of V_D when starting the MCMC chain from different values for a typical voxel

results. Similar results were obtained for both [^{11}C]diprenorphine scans indicating somewhat reproducible results (Figure 6).

Model selection Figures 5 and 6 also show the model order estimates obtained by using AIC, BIC, and Bayes factors for the data. The model selection results of AIC and BIC do not exhibit any obvious spatial structure. For both, the two-compartment model is most widely favoured. When using Bayesian model selection, the one-compartment model dominates in low activity areas. These areas are of less interest, but the findings are perhaps interesting. Identifying the parameters of a second or third compartment in areas with barely any activity is rather difficult. In the extreme case, for a voxel with no activity but noisy signals, the model can have arbitrary compartments, each of them having near zero concentration. Using Bayesian model comparison, the one-compartment model is chosen which at least favours a parsimonious representation; it could be argued that a null model should be added to the class of models under consideration to account for this case.

Within the areas of greatest interest in which a real neurological signal is expected, the two-compartments model is favoured most often. However there are more three-compartments model selected in high V_D areas. Overall the image of model order for the Bayesian analysis shows rather more spatial structure than the AIC and BIC cases although none has been imposed. It is biologically reasonable to believe that there are similar compartmental structures for voxels within the same area and different compartmental structures for voxels from difference regions. Though we don't know what the true model is, or indeed believe that there is a *true* compartmental model in this setting, the Bayesian model selection, which reveals spatial structure, is more convincing than the other two. Indeed methods which do not require the specification of a single compartmental structure for the whole brain are well known to be preferred when modelling [^{11}C]diprenorphine (Hammers et al., 2007). The different model structures can also be quantified and uncertainty attached to the estimates of the model order. Also shown in Figures 5 and 6 are the posterior model probabilities of the chosen model. For the majority of the voxels, the chosen model has a posterior probability $p(M|D) \geq 0.5$. For low V_D region, the posterior probability is much higher indicating that there is relatively high confidence that one compartment is adequate to explain what is observed in these regions but that there is a lack of strong evidence to support a particular model configuration in the case of more active regions.

Overall, the Bayesian model selection framework provides comparable parameter estimation performance with other methods such as those in Peng et al. (2008) and Jiang et al. (2009), empirically alleviating the biases associated with NNLS, but in addition yields evidence as to

the posterior probability of the chosen models. This gives valuable additional information when analysing subsequent normal or patient data. Regions where there is considerable uncertainty will require larger deviations in patient populations to establish differences, thus helping inform study designs in applications with tracers such as [^{11}C]diprenorphine. In addition, a principled way of model averaging can be trivially performed using the output from the MCMC analysis.

5. Conclusions

Throughout this work, a framework using Bayesian statistics to do parameter estimation, model comparison and model selection for PET compartmental models is illustrated. It is shown that the Bayesian estimates compare favourably with other model fitting methods. Bayesian model selection improves the MSE in the simulation case within the regime of interest. For real measured data, the Bayesian model selection gives more sensible results and allows us to directly incorporate knowledge of the compartmental system via the prior distribution.

It is of course possible to employ other computational algorithms to perform parameter estimation and model selection jointly within a simulation run, employing RJMCMC (Green, 1995), for example. In more complex problems these approaches may be more appropriate; the simple approach adopted here was adequate for the data which we have encountered thus far. The purpose of this paper is not to advocate a particular computational approach to Bayesian model selection but to show the potential gains associated with adopting a Bayesian approach to the problem in the context of PET studies in particular.

We have demonstrated that the most widely used model does not fit real PET data well and proposed a simple extension using a t-distributed noise model. This allows for the direct estimation of models even when moderate outliers are present in the data. This is very often the case with real data, for example, the delay of the input into the system is often not constant for all locations. In addition, calculating uncertainty estimates for both models and estimates are possible, something which it is inherently difficult (if not impossible) to achieve with methods based around NNLS and other point estimation techniques. This provides considerable additional information when comparing scans, and will be of particular interest when comparing normal controls verses patient groups where lesions or other pathological problems may introduce considerable differences in the uncertainty of the measures for different scans.

It would be interesting to develop methods to exploit spatial homogeneity within the brain to improve performance and produce more parsimonious inference. Further investigation into the modelling problem may also be warranted as the above demonstrate that the assumption of normally-distributed errors is not consistent with real data and with a heavier-tailed noise distribution it is not possible to obtain strong evidence in support of any one model using the type of data which is typically available. With the present modelling approach, macro parameters and other such quantities of interest could be more robustly estimated by Bayesian model averaging than by any approach based upon model selection and that is the strategy which we would recommend.

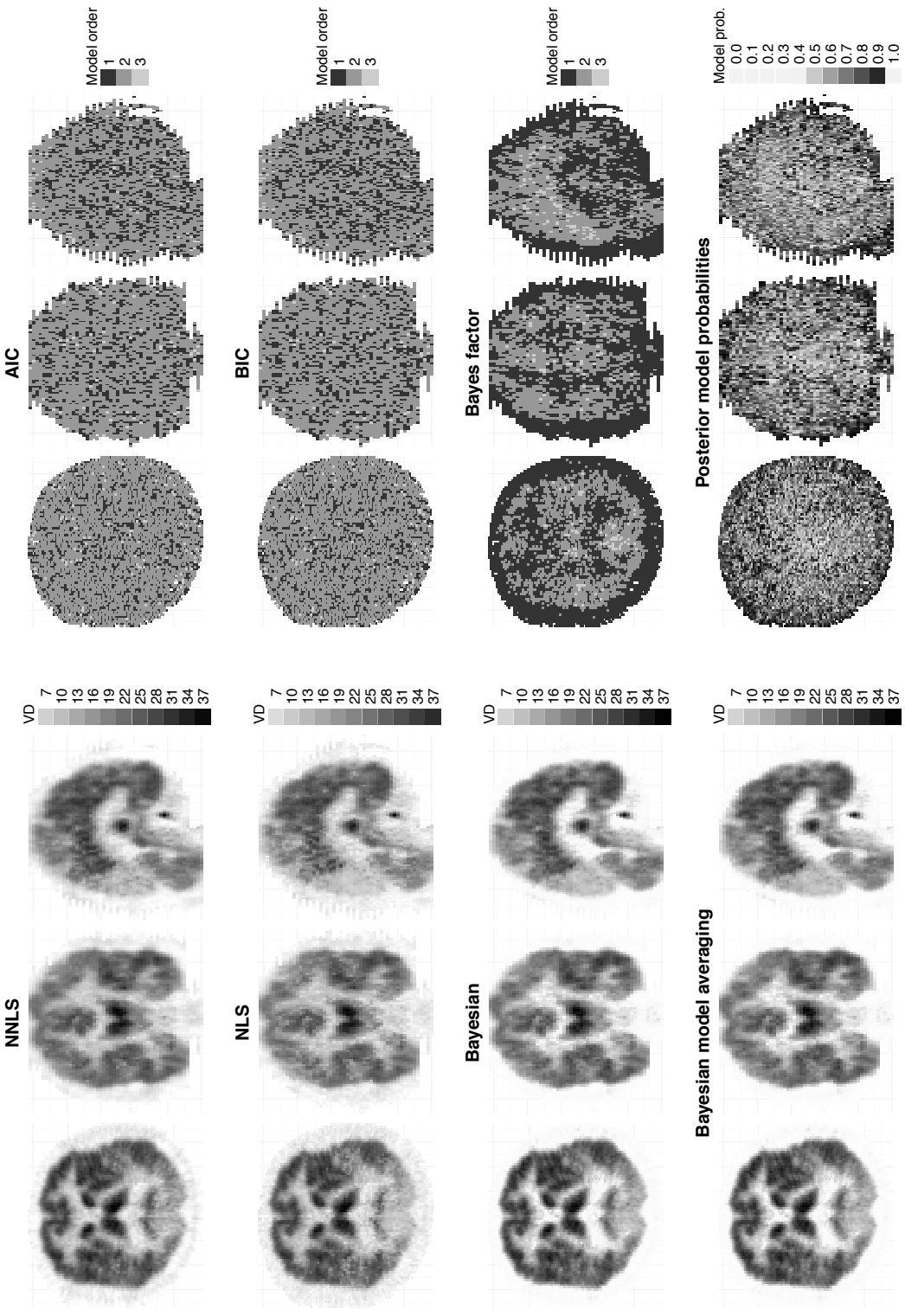

Fig. 5. Estimates for V_D and model orders the first data set, three compartments, (Jiang et al., 2009, data). Left: estimates for V_D. Right: model orders

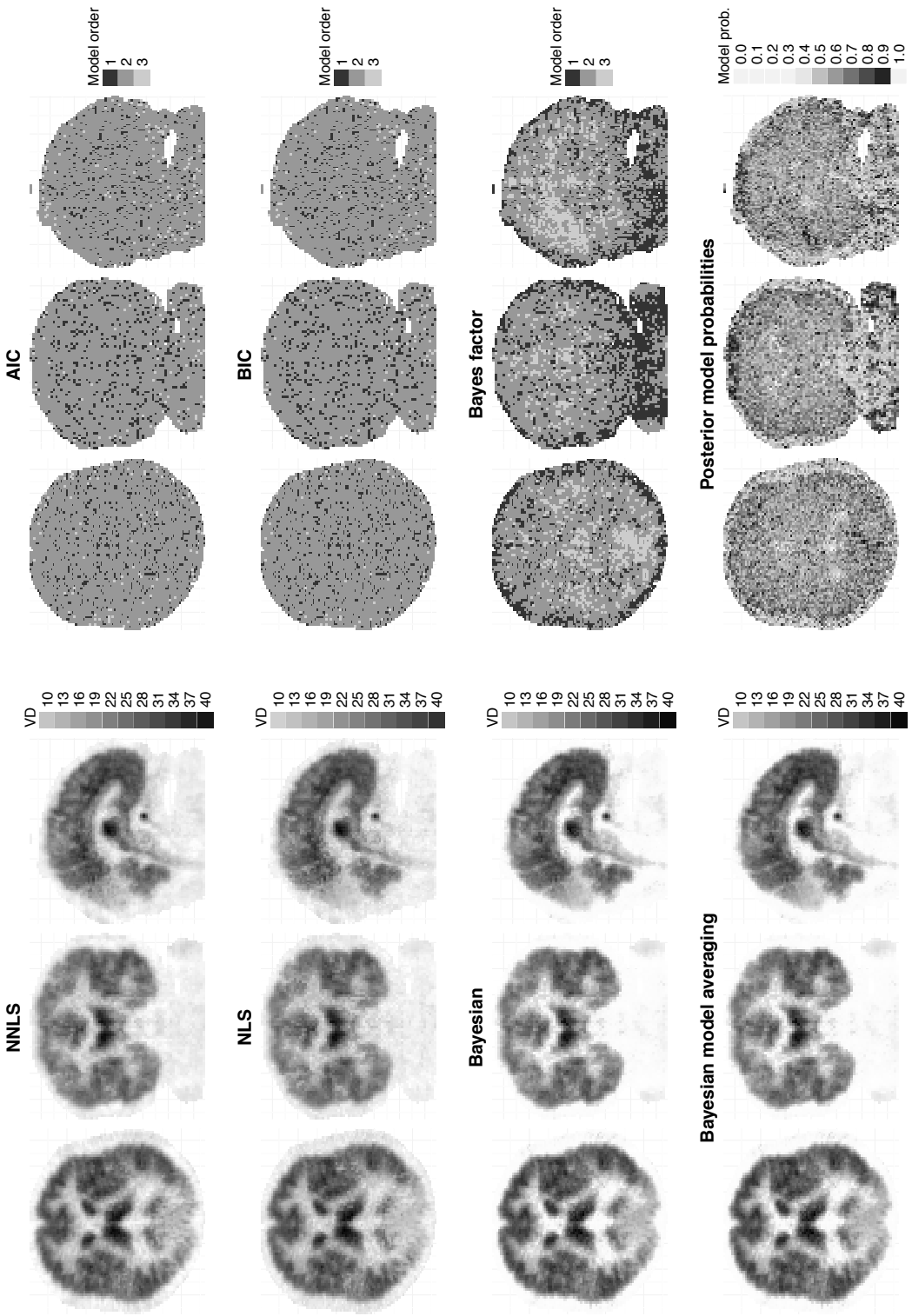

Fig. 6. Estimates for V_D and model orders the first data set, three compartments, (Peng et al., 2008, data). Left: estimates for V_D. Right: model orders

References

Akaike, H. (1973). Information theory and an extention of the maximum likelihood principle. In *Breakthroughs in Statistics, Volume I*, Springer Series in Statistics, pp. 610–624. New York, USA: Springer-Verlag.

Anderson, D. H. (1983). *Compartmental Modelling and Tracer Kinetics*, Volume 50 of *Lecture Notes in Biomathematics*. New York, USA: Springer-Verlag.

Congdon, P. (2006). *Bayesian Statistical Modelling* (3rd ed.). Wiley Series in Probability and Statistics. West Sussex, UK: John Wiley & Sons.

Cunningham, V. J. and T. Jones (1993). Spectral analysis of dynamic PET studies. *Journal of Cerebral Blood Flow and Metabolism 13*, 15–23.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (2004). *Bayesian Data Analysis*. Chapman & Hall/CRC.

Geweke, J. (1989). Bayesian inference in econometric models using Monte Carlo integration. *Econometrica: Journal of the Econometric Society 57*(6), 1317–1339.

Gibson, G. and E. Renshaw (1998). Estimating parameters in stochsatic compartmental models using Markov chain methods. *IMA Journal of Mathematics Applied in Medicine and Biology 15*, 19–40.

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

Gunn, R. N., S. R. Gunn, and V. J. Cunningham (2001, June). Positron emission tomography compartmental models. *Journal of Cerebral Blood Flow & Metabolism 21*(6), 635–52.

Hammers, A., M.-C. Asselin, F. E. Turkheimer, R. Hinz, S. Osman, G. Hotton, D. J. Brooks, J. S. Duncan, and M. J. Koepp (2007). Balancing bias, reliability, noise properties and the need for parametric maps in quantitative ligand PET: [^{11}C]diprenorphine test-retest data. *NeuroImage 38*(1), 82 – 94.

Hawkins, R. A., M. E. Phelps, and S.-C. Huang (1986, April). Effects of temporal sampling, glucose metabolic rates, and disruptions of the blood-brain barrier on the FDG model with and without a vascular compartment: studies in human brain tumors with PET. *Journal of Cerebral Blood Flow & Metabolism 6*(2), 170–183.

Hurvich, C. M. and C.-L. Tsai (1989). Regression and time series model selection in small samples. *Biometrika 76*(2), 297–307.

Innis, R. B., V. J. Cunningham, J. Delforge, M. Fujita, R. N. Gunn, J. Holden, S. Houle, S.-C. Huang, M. Ichise, H. Iida, H. Ito, Y. Kimura, R. A. Koeppe, G. M. Knudsen, J. Knuuti, A. A. Lammertsma, M. Laruelle, R. P. Maguire, M. Mintun, E. D. Morris, R. Parsey, J. Price, M. Slifstein, V. Sossi, T. Suhara, J. Votaw, D. F. Wong, and R. E. Carson (2007). Consensus nomenclature for in vivo imaging of reversibly binding radioligands. *Journal of Cerebral Blood Flow and Metabolism 27*, 1533–1539.

Jacquez, J. A. (1996). *Compartmental Analysis in Biology and Medicine* (3rd ed.). University of Michigan Press.

Jiang, C.-R., J. A. D. Aston, and J.-L. Wang (2009, August). Smoothing dynamic positron emission tomography time courses using functional principal components. *NeuroImage 47*(1), 184–93.

Jones, D. R., C. D. Perttunen, and B. E. Stuckman (1993, October). Lipschitzian optimization without the Lipschitz constant. *Journal of Optimization Theory and Applications 79*(1), 157–181.

Kass, R. E. and A. E. Raftery (1995, July). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Kinahan, P. and J. Rogers (1989). Analytic 3D image reconstruction using all detected events. *IEEE Transactions on Nuclear Science 36*(1), 964–968.

Kullback, S. and R. A. Leibler (1951). On information and sufficiency. *The Annals of Mathematical Statistics 22*(1), 79 – 86.

Mankoff, D. A., A. F. Shields, M. M. Graham, J. M. Link, J. F. Eary, and K. A. Krohn (1998). Kinetic analysis of 2-[Carbon-11]Thymidine PET imaging studies: Compartmental model and mathematical analysis. *The Journal of Nuclear Medicine 39*(6), 1043–1055.

Nelder, J. A. and R. Mead (1965). A simplex method for function minimization. *The Computer Journal 7*(4), 308–313.

Peng, J.-Y., J. A. D. Aston, R. N. Gunn, C.-Y. Liou, and J. Ashburner (2008, September). Dynamic positron emission tomography data-driven analysis using sparse Bayesian learning. *IEEE Transactions on Medical Imaging 27*(9), 1356–69.

Phelps, M. E. (2000, August). Positron emission tomography provides molecular imaging of biological processes.

Robert, C. P. (2007). *The Bayesian Choice: From Decision-Theoretic Foundations to Computational Implementation* (2nd ed.). Springer Texts in Statistics. New York, USA: Springer.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods* (2nd ed.). Springer Texts in Statistics. New York, USA: Springer.

Schmidt, K. (1999). Which linear compartmental systems can be analyzed by spectral analysis of PET output data summed over all compartments? *J Cereb Blood Flow Metab 19*(5), 560–9.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics 6*(2), 461 – 464.

Stroustrup, B. (1991). *The C++ Programming Language* (2nd ed.). Addison Wesley.

Tierney, L. (1994). Markov chains for exploring posterior distributions. *The Annals of Statistics 22*(4), 1701–1728.

Turkheimer, F. E., R. Hinz, and V. J. Cunningham (2003, April). On the undecidability among kinetic models: From model selection to model averaging. *Journal of Cerebral Blood Flow & Metabolism 23*(4), 490–498.

Wakefield, J. (1996, March). The Bayesian analysis of population pharmacokinetic models. *Journal of the American Statistical Association 91*(433), 62–75.