Exponential Model Selection (in NMR) Using Bayesian Probability Theory

G. LARRY BRETTHORST,[1] WILLIAM C. HUTTON,[1] JOEL R. GARBOW,[1,2] JOSEPH J.H. ACKERMAN[1–3]

[1] *Department of Radiology, Washington University, One Brookings Drive, St. Louis, MO 63130, USA*
[2] *Department of Chemistry, Washington University, One Brookings Drive, St. Louis, MO 63130, USA*
[3] *Department of Internal Medicine, Washington University, One Brookings Drive, St. Louis, MO 63130, USA*

ABSTRACT: In a companion article in this issue, parameter estimation using exponential models was addressed when the form of the model is known (i.e., when the number of exponentials and whether a constant offset is present are known). In this article, we apply Bayesian probability theory to the problem of determining the functional form of the model. The calculations are implemented using Markov chain Monte Carlo with simulated annealing to draw samples from the joint posterior probability for the parameters and the functional form of the model. Monte Carlo integration is then used to approximate the marginal posterior probabilities for all the parameters, including the number of exponentials and whether a constant offset is present. Examples using empirical data are given to illustrate the calculations.
 Concepts Magn Reson Part A 27A: 64–72, 2005

KEY WORDS: exponential data analysis; Bayesian probability theory

INTRODUCTION

In a companion article in this issue (*1*), the parameter estimation problem given an exponential model was addressed using Bayesian probability theory. In that article, the number of exponentials and whether a constant offset is present were assumed to be known. As demonstrated in (*1*), if the exponential model is not in one-to-one correspondence with the data, the parameter estimates may significantly disagree with the estimates obtained using the correct model. Consequently, it is important for the model to correspond exactly to the data. In this article, we apply Bayesian probability theory to the problem of determining the number of exponentials and whether a constant offset is present.

The set of models considered has four different functional forms. The most complex model considered is

$$d_i = C + \sum_{j=1}^{m} A_j \exp\{-\alpha_j t_i\} + n_i \qquad [1]$$

Received 14 June 2005; revised 24 August 2005; accepted 24 August 2005

Correspondence to: G. Larry Bretthorst; E-mail: gbretthorst@wustl.edu

Concepts in Magnetic Resonance Part A, Vol. 27A(2) 64–72 (2005)

Published online in Wiley InterScience (www.interscience.wiley.com). DOI 10.1002/cmr.a.20042

where d_i is the data item sampled at t_i, C is the constant offset, A_j and α_j are the amplitude and decay rate constant of the jth exponential, n_i represents the noise or measurement error, and m is the number of exponentials. The second form is

$$d_i = \sum_{j=1}^{m} A_j \exp\{-\alpha_j t_i\} + n_i \qquad [2]$$

where this model differs from the first in that the constant offset is not present. Finally, the last two functional forms are the two degenerate cases: no signal,

$$d_i = n_i \qquad [3]$$

and constant offset only

$$d_i = C + n_i. \qquad [4]$$

THEORY

In Bayesian probability theory, model selection problems are solved by computing the posterior probability for the model given the data and the prior information. To perform this calculation, the model must be part of the inference problem. Equations [1–4] are the four different functional forms considered, but those equations do not by themselves designate a model. For example, Eq. [1] is a set of m different models. To indicate a model, we define a model indicator u. If we designate the no-signal model as $u = 0$ and the constant-offset model as $u = 1$, then a natural progression would be to designate a single exponential with no constant as $u = 2$, a single exponential with a constant as $u = 3$, a biexponential with no constant as $u = 4$, and so on. Thus, the model indicator takes on values $u \in \{0, 1, \ldots, U_{max}\}$, where U_{max} is the maximum value of the model indicator, which we assume to be known. In the program that implements this calculation, the maximum is set to $U_{max} = 9$, four exponentials plus a constant. The reason we do not go higher is that distinguishing four exponentials plus a constant already requires hundreds of data values with signal-to-noise ratios of many thousands.

As noted, in Bayesian probability theory the model selection problem is solved by computing the posterior probability for the model. In this case, that means computing the posterior probability for the model indicator given the data and the prior information. This posterior probability is represented symbolically by $P(u|D\,I)$, where D represents all of the data and I represents all of the prior information. Arguments appearing inside a formal probability symbol are hypotheses. For example, when $u = 5$, the hypothesis we are making inferences about is of the form, "When the data were taken, the signal was biexponential plus a constant." Similarly, if $u = 2$, the hypotheses is of the form, "When the data were taken, the signal was single exponential without a constant." Eventually, we reach the point where we assign numerical values to represent probabilities. Then parameters appearing in these assignments will be numbers that designate hypotheses. A notation could be developed that distinguishes these two different usages, but such notation would be clumsy at best and the distinction between a hypothesis and a parameter is usually clear from the context of the equations.

To compute the posterior probability for the model indicator, one applies Bayes' theorem (*2*, *3*) to obtain

$$P(u|DI) = \frac{P(u|I)\,P(D|uI)}{P(D|I)}. \qquad [5]$$

The three terms on the right-hand side of this equation are the prior probability for the model indicator, $P(u|I)$, the direct probability for the data given the model indicator, $P(D|uI)$, and the probability for the data given the prior information, $P(D|I)$.

The probability for the data given the prior information is computed from the joint probability for the data and the model indicator given only the prior information, $P(uD|I)$, by application of the sum rule:

$$P(D|I) = \sum_{u=0}^{U_{max}} P(uD|I) = \sum_{u=0}^{U_{max}} P(u|I)\,P(D|uI). \qquad [6]$$

Examining Eq. [6], we see that this probability is exactly the quantity needed to ensure that Eq. [5] is normalized. If we normalize Eq. [5] at the end of the calculation, the normalization constant can be dropped giving

$$P(u|DI) \propto P(u|I)\,P(D|uI). \qquad [7]$$

The probability for the data given the model indicator and the prior information, $P(D|uI)$, depends only on the data and u; yet any given model has parameters (i.e., hypotheses) associated with it. Even the degenerate no-signal model, $u = 0$, has a standard deviation associated with it. Consequently, $P(D|uI)$ must be a marginal probability. Marginal probabilities are probabilities that have had the dependence on one or more

hypotheses removed using the sum rule of probability theory. If we designate the parameters in the indicated model as Ψ_u, the probability for the data given the model indicator, $P(D|uI)$, may be computed from the joint probability for the data and the parameters, $P(\Psi_u D|uI)$, by application of the sum rule:

$$P(D|uI) = \int d\Psi_u P(\Psi_u D|uI) \qquad [8]$$

where the integral is over all parameters in the indicated model. If $u = 5$, the biexponential plus a constant model, then $\Psi_u \equiv \{C, \alpha_1, \alpha_2, A_1, A_2, \sigma\}$; where σ is the standard deviation of the prior probability for the noise. Consequently, when $u = 5$, the integral is six-dimensional. Note that this equation involves the dual usage of hypotheses and parameters; inside $P(\Psi_u D|uI)$, the Ψ_u designate hypotheses, whereas outside, in the integral, they designate numbers.

The integrand in Eq. [8] may be factored using the product rule of probability theory to obtain

$$P(D|uI) = \int d\Psi_u P(\Psi_u|uI)\, P(D|\Psi_u uI). \qquad [9]$$

where $P(D|\Psi_u\, uI)$ is the direct probability for the data given the parameters and the model indicator, and $P(\Psi_u|uI)$ is the prior probability for the parameters given the model indicator.

Substituting the probability for the data given the model indicator, Eq. [9], into the posterior probability for the model indicator, Eq. [7], gives

$$P(u|DI) \propto P(u|I) \int d\Psi_u P(\Psi_u|uI)\, P(D|\Psi_u uI). \qquad [10]$$

The integrand in Eq. [10] is the prior probability for the parameters given the model indicator times the direct probability for the data, given the parameters and the model indicator. This integrand *is* the posterior probability for the parameters given the model indicator, and the calculation of this probability was addressed in (*1*). The details of factoring and assigning these probabilities are not repeated here, and we use the results from (*1*).

First, the prior probability for the model indicator, $P(u|\mathrm{I})$, must be assigned. Following (*1*), we assign this prior probability using a bounded Gaussian of the form:

$$P(u|I) \propto \begin{cases} \exp\left\{-\dfrac{(4-u)^2}{2\times 3^2}\right\} & \text{if } 0 \le u \le 9 \\ 0 & \text{otherwise,} \end{cases} \qquad [11]$$

for which the mean model indicator is 4 (biexponential with no constant), the maximum is 9 (four exponentials plus a constant), and the standard deviation of this prior probability is 3. Consequently, this prior expresses a slight preference for biexponentials and indicates that we think it about one fourth as likely that the number of exponentials is as high as 4 or as low as 0. One could easily modify this prior to allow models with higher orders but as noted, exceedingly high-quality data would be required to justify such an analysis.

DISCUSSION

One could evaluate the integral in Eq. [10] for each value of the model indicator. Indeed, the Markov chain Monte Carlo simulation that implements the calculation in (*1*) does just that for each model processed. As a result, one can do simple model comparisons using the algorithm described in (*1*) . However, the process is cumbersome and inefficient. It is better to treat the model indicator as a parameter and sample all the parameters, including the model indicator, using Markov chain Monte Carlo to approximate the Bayesian posterior probability. The integrals in Eq. [10] are then evaluated using Monte Carlo integration. When this is done, the target distribution of the Markov chain is the joint posterior probability for all of the parameters, including the model indicator, Eq. [10], without the integrals.

The calculation is implemented using Markov chain Monte Carlo with simulated annealing. In simulated annealing, one introduces an annealing parameter to facilitate exploring the parameter space. The annealing parameter, β, is introduced around the joint posterior probability for the parameters given the model indicator:

$$P(u\Psi_u|DI) \propto P(u|I)[P(\Psi_u|uI)\, P(D|\Psi uI)]^{\beta}. \qquad [12]$$

The inverse of β is often referred to as an annealing temperature.

To implement this calculation, multiple Markov chain Monte Carlo simulations are run in parallel using Eq. [12] as the target distribution. The simulations are initialized with samples from the prior probabilities. To initialize a simulation, a random sample is drawn from the prior probability for the model

Table 1 The Distribution of Model Indicators, *u*, as a Function of β

β	$\langle(\log P(D\|I)\rangle_\beta$	0	0 + *C*	1	1 + *C*	2	2 + *C*	3	3 + *C*	4	4 + *C*
0.00	−29.6280	1	12	9	19	16	13	18	6	3	3
0.10	−26.4830	10	9	19	19	16	15	8	3	1	0
0.20	−21.0090	3	4	22	27	26	9	7	2	0	0
0.30	−9.9270	1	2	43	15	20	10	7	1	1	0
0.40	−7.0259	0	0	40	25	25	6	3	1	0	0
0.50	−4.8198	0	0	55	21	16	6	2	0	0	0
0.60	−2.6279	0	0	39	16	32	12	1	0	0	0
0.70	−0.8594	0	0	26	14	49	10	1	0	0	0
0.80	−0.5635	0	0	26	13	50	11	0	0	0	0
0.90	2.7206	0	0	3	0	91	6	0	0	0	0
1.00	2.6850	0	0	0	0	86	14	0	0	0	0

indicator. For example, if this model indicator was for a two exponential plus a constant model, then samples for the two decay rate constants, the two amplitudes, and the constant offset are drawn randomly from their respective prior probabilities. The six parameters (C, α_1, α_2, A_1, A_2, u) are used as the starting values for one of the Markov chain Monte Carlo simulations. The remaining simulations are initialized in a similar manner, but because the parameters are initialized randomly, all of the Markov chain Monte Carlo simulations start out at different locations in the parameter space.

Table 1 illustrates the model selection process as a function of the annealing parameter using 100 independent simulations. The first column is the current value of β. The second column is the logarithm of the likelihood averaged over the 100 simulations, and the remaining columns are the number of simulations having a given model indicator, u. When the annealing process begins, $\beta = 0$, and the simulations are distributed according to the prior probability for the model indicator. This is can be seen in the first row of Table 1. For example, the 16 under the "2" heading means there were 16 simulations having two exponentials with no constant. Because we use a slightly informative prior probability for the model indicator, most of the simulations cluster around the one, two, and three exponential models.

A Metropolis-Hastings algorithm (*4*, *5*) was used to implement the Markov chain Monte Carlo simulation. In a Metropolis-Hastings algorithm, one proposes a new value for one of the parameters, calculates Eq. [12] using the new parameter, and then accepts the new parameter if the probability increases. If the probability decreases, the new parameter may still be accepted. If the odds ratio (new probability/previous probability) is greater than a randomly drawn uniform deviate, the new parameter value is accepted. Otherwise, the new parameter is rejected and the simulation, the parameters, remain unchanged.

The proposed value of a new parameter is generated using a Gaussian random number generator centered on the old value of the parameter. The standard deviation of this Gaussian is adjusted between annealing steps to keep the acceptance rates around 25% for a given type of parameter. Acceptance rates greater than about 30% imply the simulations are not adequately exploring the parameter space, whereas acceptance rates lower than about 20% imply the simulations are not sampling the posterior probability adequately.

Proposals for the model indicator are more difficult because when the model indicator changes, all of the parameters change. We propose a change in the model indicator by increasing or decreasing the model indicator randomly using a Gaussian random number generator. The parameters associated with the new model indicator are proposed by cloning an existing model, if possible, or by sampling the prior if no model of the proposed type exists. In either case, the proposed model parameters, excluding the model indicator, are simulated until the Metropolis-Hastings algorithm reaches equilibrium at the current value of the annealing parameter. The proposed model and all of its parameters are then accepted or rejected using the criteria described in the previous paragraphs.

At a fixed value of the annealing parameter, each simulation is processed in parallel and completely independently of the other simulations. Each parameter in each simulation is processed at least 25 times. After the simulations reach equilibrium at a given value of the annealing parameter, one line of the report shown in Table 1 is generated. The annealing parameter is then increased. The simulations are sorted by the joint posterior probability for the parameters (computed at the new value of the annealing parameter) and a number, 3%, of the simulations

having the lowest posterior probabilities are replaced by higher probability simulations. This step, removing low probability simulations, is essential to make sure simulations do not get stuck in a local maximum of joint posterior probability, Eq. [12]. Finally, the entire process is repeated until the annealing parameter reaches one.

When the annealing process is finished, $\beta = 1$, the simulations have distributed themselves into the joint posterior probability for the parameters and the model indicator. To understand how the simulations do this, think of the simulations as a swarm looking for the maximum of the joint posterior probability. When the simulations starts, $\beta = 0$, and the simulations distribute themselves over the allowed parameter and model space. For small values of the annealing parameter, the simulations easily transition between different models. This effect is illustrated in Table 1 because at small annealing parameter values the number of simulations having a given model indicator keeps changing.

As the annealing parameter increases, the joint prior probability for the parameters and the direct probability for the data given the parameters becomes increasingly important. Suppose a proposed model has two exponentials plus a constant, and the data have two exponentials. The signal is already fit down to the noise by the two exponential model, so the constant fits the noise. The likelihood function, essentially $P(D|\Psi_u uI)$, cannot increase much because the signal is already fit to the noise by the embedded two-exponential model. However, the prior probability for the constant decreases the joint posterior probability, so transitions to this model are not favored. As the annealing parameter increases, the swarm spends less and less time searching this model because the peak in the joint posterior probability is not there. This effect is again illustrated in Table 1 because complicated models are abandoned as the annealing parameter increases; the extra parameters do not improve the fit to the data, and the prior probabilities rule the complicated models out.

Similarly, if a model is proposed that has fewer exponentials than needed to model the data, the likelihood will not be able to fit the data as well as a two-exponential model. In this case, transitions to the simpler model will almost certainly stop as the annealing parameter increases. Again, the swarm stops visiting models that cannot fit the data down to the noise. What one typically sees is a sharp cutoff below which the models cannot fit the data. Note that the cutoff is pronounced in Table 1. By the time the annealing parameter has reached 0.9, few simulations visit these low-probability models.

The results of Markov chain Monte Carlo simulations are only meaningful if those simulations have converged. In this context, convergence means that the simulations have reached a stationary point of the distribution. At a stationary point, the mean and standard deviations of the samples become time independent. We assess convergence visually by plotting the logarithm of the posterior probability for each simulation as a function of time. Because each simulation independently samples the posterior probability, its trajectory should be a random walk. Thus, when converged, a plot of all the trajectories is stationary, random, and exhibits no discernible patterns.

Within a given model, the distribution of simulations reflects the joint posterior probability for the parameters given the model indicator. Across models, the distribution of simulations reflects the joint posterior probability for the model indicator. The program uses all of the samples from all of the high-probability models to output parameter estimates and histograms. The histograms are used for display purposes only and provide a visual representation of the marginal posterior probability for each parameter given the model indicator. The parameter estimates are in the form of a mean and standard deviation of the samples from a given model indicator. The program also outputs the parameters from the simulation that had the maximum joint posterior probability given the model indicator. Because these estimates are computed from the samples drawn from the joint posterior probability for the parameter given the model indicator and the data, they are identical (to within the uncertainty in the estimate) to those obtained using the calculations presented in (*1*).

For more information on Markov chain Monte Carlo, see (*6*, *7*), and for more details on how to implement model selection within simulated annealing, see (*8*, *9*).

EXAMPLES

In (*1*), we reported parameter estimates from a ^{23}Na transverse relaxation study on rat brain in vivo. Based on literature reports, we assumed the model for the ^{23}Na relaxation data was either single or biexponential. A qualitative comparison of the residuals from these two models showed a systematic artifact in the single exponential model, whereas the biexponential model had no such artifact. This qualitative comparison supports the hypotheses that the data are biexponential. Now that we have addressed the unknown number of exponentials problem, we can report a quantitative analysis of these data.

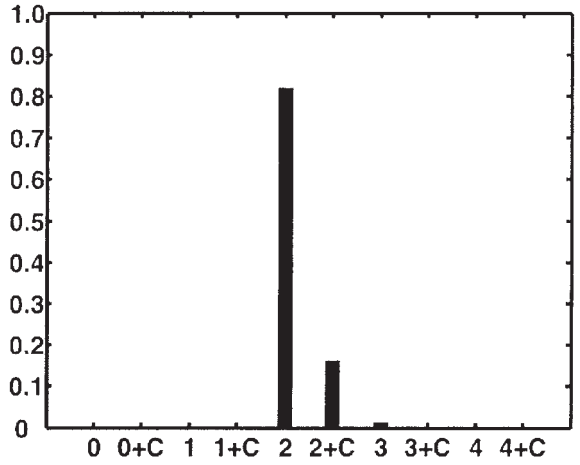

Figure 1 The posterior probability for the model indicator was computed using the ^{23}Na NMR relaxation data described in (*1*). These data were also used in generating Table 1. The posterior probability has a peak at two exponentials. There are two smaller peaks at two exponentials plus a constant and at three exponentials (the latter about 0.01). Bayesian analysis indicates the ^{23}Na NMR relaxation data are biexponential.

Using Markov chain Monte Carlo and the ^{23}Na relaxation data, the joint posterior probability for all of the parameters, including the model indicator, was sampled. If Monte Carlo integration is used to remove the dependence on all of the parameters except the model indicator, the resulting samples are taken from the posterior probability distribution for the model indicator. We have plotted this posterior probability as a series of impulses in Fig. 1. Note that this posterior probability is, up to a normalization constant, virtually identical to Table 1. The reason for this is that the ^{23}Na data were also used in generating Table 1. Below two exponentials, the probability distribution is 0, indicating that neither a 1 nor a 1 exponential plus a constant model correctly accounts for the structure in these data. The probability distribution has a peak at two exponentials, again giving support to the hypotheses that the data are biexponential. Additionally, this probability distribution has small peaks at two exponentials plus a constant and at three exponentials. All models more complex that three exponentials without a constant have zero probability; i.e., no simulations survived for these models. Given that the total probability for biexponentials models is about 0.99, the analysis strongly indicates that the ^{23}Na data are biexponential.

As a second example, we use data taken from MRI eddy current measurements (Rolf Gruetter, private communication; see (*10*)). On MR-imaging scanners, eddy currents are induced by rapidly switched magnetic field gradients, and compensation is required to minimize the image artifacts resulting from these eddy currents. Eddy current compensation alters gradient input waveforms via multiple preemphasis exponential time constants to achieve a more ideal output wave form. Typically, MRI equipment vendors provide up to four exponential time constants and amplitudes on each gradient channel (X, Y, Z) and on a B_0 correction coil for eddy current compensation. The challenge is to determine the number of exponential components (e.g., mono-, bi-, tri-) and the parameters (time constants and amplitudes) that define these components.

To first approximation (ignoring cross-terms between gradient channels and nonlinear effects), the eddy currents from a given gradient pulse produce: (i) a time-dependent shift in B_0 that is independent of coordinate position and (ii) a time-dependent shift in B_0 that is linear in coordinate position along the gradient axis (i.e., a time-dependent gradient G). Thus, for each of the three gradient channels there are two data sets that describe the eddy currents following a gradient pulse: one that quantifies the time dependence of B_0 and one that quantifies the time dependence of G. The time dependence of B_0 can conveniently be expressed as a changing resonance frequency ($\omega = \gamma B_0$). The time dependence of G can conveniently be expressed as a percentage of the applied gradient strength. Absolute values are not important and the data has been normalized for display and computation purposes.

Here, we describe the time-dependent eddy current components induced by a gradient pulse from the *i*th gradient channel (i = X, Y, Z) as $B_{0i}(t)$ and $G_i(t)$. For example, following a gradient pulse along the X axis, eddy currents described by $B_{0X}(t)$ and $G_X(t)$ are present. For conciseness, we drop the explicit time dependence from the notation. Figure 2 is a plot of three of six representative eddy current data sets. Figure 2(A) shows both B_{0Y} and G_Y data, whereas B_{0X} data are shown in Fig. 2(B). Note that abscissa and ordinate scales are different for Fig. 2(A) and Fig. 2(B).

Table 2 summarizes the Bayesian analyses of these six eddy current data sets. The parameter estimates, mean $\pm$ standard deviation, are listed in Table 2 for all high probability models. The program that implements the Bayesian calculation does the calculations using decay rate constants, and it is decay rate constants that are shown in Table 2. However, the program also outputs the decay time constants. In addition to the parameter estimates for the decay rate constants, amplitudes, and constant offsets, Table 2 also shows the type of model under the heading *m*, the

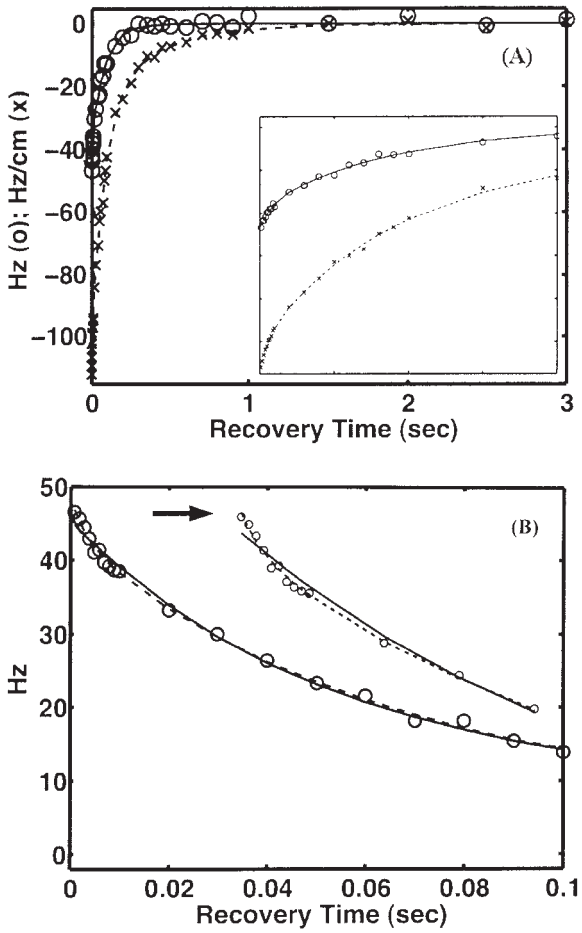

Figure 2 Panel (A) is a plot of the B_{0Y} (o) and G_Y (x) eddy current data. The B_{0Y} data are modeled as two exponentials (solid line), while the G_Y data are modeled as three exponentials (dashed line). The inset is an expansion of the first 0.2 s of these data. Panel (B) is a plot of the B_{0X} data, note the different scales. The inset (*arrow*) is an expansion of the first 30 ms of these data. Probability theory indicated two high-probability models for the B_{0X} data. These models, two and three exponential, are shown as the solid line and dashed line respectively.

number of exponentials in the model, Eqs. [1–4]. We have augmented the number of exponentials with a "$+c$," when necessary, to indicate that a constant offset is included in the model.

In three of the six data sets, B_Z, G_X and G_Y, probability theory strongly prefers a three exponential model. In these three cases, the posterior probability for a three exponential model exceeds 0.90. When present, the third exponential has a small amplitude (A = 5–19%) and decays rapidly ($\alpha \sim$ 200–330 sec^{-1}). Given the uncertainty in the third decay rate constant, it is possible that this third component is the same in all three data sets. In each of these three data sets, the probability that a constant is present is approximately 0.1.

In Fig. 2(A) we have plotted the G_Y data and model. This model, computed from the parameters of the Markov chain Monte Carlo simulation having maximum joint posterior probability for the parameters, is displayed as the dashed lines. As the inset shows, this model does an excellent job of representing these data.

In two of the six data sets, B_{0Y} and G_Z, probability theory prefers a two exponential model. We have plotted one of these data sets, B_{0Y}, in Fig. 2(A). In a manner similar to the three exponential case we also generated a model of these two exponential data and have plotted this model as the solid line in Fig. 2(A). As the inset shows, this model also does an excellent job of representing these data.

Finally, one data set, B_{0X}, gave a mixed result: there was a probability of 0.38 that the number of exponentials was two, and a probability of 0.47 that the number of exponentials was three. The remaining probability (0.15) was split among models having two or three exponentials with constant offsets. We have plotted these data sets in Fig. 2(B) as the "o" characters. We have also plotted the models generated from the two (solid line) and three (dashed line) exponen-

Table 2 Bayesian Model Selection and Parameter Estimates for Empirical Eddy Current Data

Data Set	α_1	A_1	α_2	A_2	α_3	A_3	C	$P(u\|DI)$	m
B_{0Z}	2.3 ± 0.18	−38 ± 3.2	14 ± 1	−79 ± 2.7	365 ± 113	−18 ± 2.5	—	0.77	3
	2.5 ± 0.2	−39 ± 3.2	15 ± 1	−78 ± 2.8	376 ± 108	−17 ± 2.3	−1 ± 0.3	0.21	3+c
B_{0X}	3.1 ± 0.5	13 ± 2.4	20 ± 2	31 ± 2.2	—	—	—	0.38	2
	2.8 ± 0.5	12 ± 2	18 ± 1.6	32 ± 2	345 ± 108	5 ± 1.6	—	0.47	3
B_{0Y}	11.7 ± 0.6	−37 ± 1.4	162 ± 52	−10 ± 1.5	—	—	—	0.99	2
G_Z	2.8 ± 0.3	−29 ± 3.6	12.5 ± 0.6	−80 ± 3.5	—	—	—	0.93	2
G_X	3.4 ± 0.34	−33 ± 3.9	17 ± 1.6	−52 ± 3.6	370 ± 108	−11 ± 2.5	—	0.81	3
	3.7 ± 0.4	−34 ± 4.2	17 ± 1.7	−51 ± 3.8	377 ± 104	−12 ± 2.5	−0.4 ± 0.2	0.10	3+c
G_Y	2.8 ± 0.8	−27 ± 2.7	12 ± 0.5	−73 ± 2.5	219 ± 45	−15 ± 1.4	—	0.89	3
	2.8 ± 0.2	−27 ± 2.5	12 ± 0.5	−74 ± 2.3	218 ± 44	−15 ± 1.4	0.03 ± 0.02	0.11	3+c

tial simulations that had maximum joint posterior probability. Though the three exponential model does fit the data better, the data are not of sufficient quality for probability theory to select this model conclusively.

SUMMARY AND CONCLUSION

Bayesian probability theory is a general theory of inference that identifies what needs to be calculated in a given problem. However, it does not specify how to do the calculations. Markov chain Monte Carlo is a general method for obtaining samples from Bayesian posterior probability distributions. Bayesian probability theory, combined with Markov chain Monte Carlo, provides a general and robust means to analyze data modeled as a sum of exponentials and provides a means to do model selection calculations easily. In addition, running multiple Markov chain Monte Carlo simulations simultaneously allows us to take advantage of parallel computing architectures to design fast and efficient computer code to implement the Bayesian calculations. We have implemented these Bayesian calculations on single and parallel computational architectures, and the software is fully integrated into a commercial NMR data analysis package (VnmrJ, Varian NMR Systems, Palo Alto, CA). Both the VnmrJ implementation and stand-alone executables for manual use are available for free download at http://BayesianAnalysis.wustl.edu.

ACKNOWLEDGMENTS

We thank Dr. Jeffrey J. Neil for his encouragement, support, and comments. We thank James Goodman for supplying the ^{23}Na relaxation data and Dr. Rolf Gruetter (University of Minnesota) for providing the eddy current data. This work was supported by a contract with Varian NMR Systems, Palo Alto, CA; by the Small Animal Imaging Resource Program (SAIRP) of the National Cancer Institute, grant R24 CA83060; and by the National Institute of Neurological Disorders and Stroke, grants NS35912 and NS41519.

REFERENCES

1. Bretthorst GL, Hutton WC, Garbow JR, Ackerman JJH. 2005. Exponential parameter estimation (in NMR) using Bayesian probability theory. Concepts Magn Reson Part A 27A:55–63.
2. Bayes Rev. T. 1763. An essay toward solving a problem in the doctrine of chances. Philos Trans R Soc London 53:370–418; reprinted in 1958. Biometrika 45:293–315; and 1963. Facsimiles of two papers by Bayes with commentary by Deming WE. New York: Hafner.
3. Jaynes ET. 2003. Probability theory: the logic of science. Bretthorst GL, editor. Cambridge: Cambridge University Press.
4. Metropolis N, Rosenbluth AW, Rosenbluth MN, Teller AH, Teller E. 1953. Equations of state calculations by fast computing machines, J Chem Phys 21:1087–1091.
5. Hastings WK. 1970. Monte Carlo sampling methods using Markov chains and their applications. Biometrika 57:97–109.
6. Gilks WR, Richardson S, Spiegelhalter DJ. 1996. Markov chain Monte Carlo in practice. London: Chapman & Hall.
7. Neal RM. 1993. Probabilistic inference using Markov chain Monte Carlo methods. Technical Report CRG-TR-93-1. University of Toronto Department of Computer Science, Toronto, Ontario.
8. Skilling J. 1998. Probabilistic data analysis: an introductory guide. J Micros 190:28–36.
9. Goggans PM, Ying C. 2004. Using thermodynamic integration to calculate the posterior probability in Bayesian model selection problems. In: Rychert J, Erickson G, Smith CR, editors. Bayesian inference and methods in science and engineering. College Park, MD: American Institute of Physics p 59–66.
10. Terpstra M, Andersen PM, Gruetter R. 1998. Localized eddy current compensation using quantitative field mapping. J Magn Reson 131:139–143.

BIOGRAPHIES

G. Larry Bretthorst is a senior scientist at Washington University. He has written extensively on parameter estimation and model selection using Bayesian probability theory, including a book titled *Bayesian Spectrum Analysis and Parameter Estimation* and more recently as editor of *Probability Theory: The Logic of Science*.

William C. Hutton received his B.A. in chemistry from Maryville College in Tennessee in 1971. In 1972, he joined the staff at the University of Virginia in the Department of Chemistry. At Virginia, he conducted research in phospholipid membrane biophysics, metal binding in biological molecules, and natural product structure elucidation. In 1983, Mr. Hutton joined Monsanto Company. At Monsanto he conducted research in life sciences NMR and laboratory automation. In 1994 he was appointed a Monsanto Fellow. Mr. Hutton joined the Washington University Medical School in 2003, where he is currently a research associate.

Joel R. Garbow, Ph.D., is a senior scientist in chemistry and radiology working in the Biomedical MR Laboratory at Washington University in St. Louis. Dr. Garbow received his undergraduate degree from the University of Illinois, Urbana-Champaign, and his Ph.D. in chemistry from the University of California, Berkeley. He joined Washington University in 2000 following a 17-year career at Monsanto Company where he directed a solid-state NMR spectroscopy-based research program. At Washington University, his current research interests include the development of MR imaging techniques and their application to the study of small-animal models of cancer.

Joseph J. H. Ackerman, Ph.D., is the William Greenleaf Eliot Professor of Chemistry and chair of the Department of Chemistry at Washington University in St. Louis where he also holds appointments as professor of radiology and research professor of chemistry in medicine. Dr. Ackerman and collaborators pursue research on the development and application of magnetic resonance spectroscopy and imaging for study of intact biological systems, from single cells to genetically engineered mice to man. A primary focus is the use of water diffusion-sensitive MR methods to probe tissue architecture and microstructure at the micron length scale, far less than the actual voxel resolution of the image itself. Dr. Ackerman is a Fellow and Gold Medal winner of the (International) Society for Magnetic Resonance in Medicine.