Bayesian Inference and Maximum Entropy in the Physical Sciences, R. Fischer et al *ed*, pp 3-14, AIP 2004.

Characterizing Water Diffusion In Fixed Baboon Brain[1]

G. Larry Bretthorst, PhD[*], Christopher D. Kroenke, PhD[†] and Jeffrey J. Neil, MD, PhD[**]

[*]*Biomedical Magnetic Resonance Laboratory, Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO 63110, (gbretthorst@wustl.edu)*
[†]*Biomedical Magnetic Resonance Laboratory, Mallinckrodt Institute of Radiology, Washington University, St. Louis, MO 63110, (kroenke@wustl.edu)*
[**]*Pediatric Neurology, Washington University School of Medicine, St. Louis, MO 63110, (neil@wustl.edu)*

Abstract. In the Biomedical Medical Research laboratory in St. Louis Missouri there is an ongoing project to characterize water diffusion in fixed baboon brain using diffusion weighted magnetic resonance imaging as a means of monitoring development throughout gestation. Magnetic resonance images can be made sensitive to diffusion by applying magnetic field gradients during the pulse sequence. Results from the analysis of diffusion weighted magnetic resonance images using a full diffusion tensor model do not fit the data well. The estimated standard deviation of the noise exhibit structures corresponding to known baboon brain anatomy. However, the diffusion tensor plus a constant model overfits the data: the residuals in the brain are smaller than in regions where there is no signal. Consequently, the full diffusion tensor plus a constant model has too many parameters and needs to be simplified. This model can be simplified by imposing axial symmetry on the diffusion tensor. There are three axially symmetric diffusion tensor models, prolate, oblate, and isotropic; and two other models, no signal and full diffusion tensor, that could characterize the diffusion weighted images. These five models may or may not have a constant offset, giving 10 total models that potentially describe the diffusion process. In this paper the Bayesian calculations needed to select which of the 10 models best characterizes the diffusion data are presented. The various outputs from the analysis are illustrated using one of our baboon brain data sets.

INTRODUCTION

The problem we would like to solve is to determine which of the 10 models mentioned in the abstract best describes the diffusion process given our baboon brain data. Each model represents the three-dimensional displacement distribution of the water in a given pixel. The data consists of a series of diffusion weighted ^{1}H magnetic resonance images. Figure 1 is the first 12 of the 45 diffusion weighted images used in this paper. Each image has been made sensitive to the diffusion of water (^1H_2O) by the application of magnetic field gradients. Magnetic field gradients sensitive the images to diffusion along the direction of the gradients. The stronger the gradients the greater the signal attenuation due to water motion. Gradients are vector quantities having both magnitude and direction. In generating the gradient vectors, we used spherical coordinates. The magnitude of the diffusion sensitizing gradients was uniformly increased for each successive diffusion weighted image. However, the two angles associated with the gradient direction were

[1] in Bayesian Inference and Maximum Entropy Methods in Science and Engineering, Rainer Fischer, Roland Preuss and Udo von Toussaint *eds.*, AIP conference proceedings number **735**, pp. 3-15, 2004.

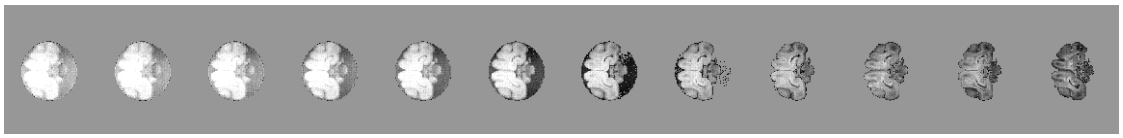

FIGURE 1. In diffusion weighted images, the images are made sensitive to diffusion of water by application of magnetic field gradients. These gradients sensitive the experiment to motion along the direction of the gradients. The gradients starts at zero, left image, and increases going to the right. For the right-most image the signal from the rapidly-diffusing water in formalin, a liquid bathing the brain, has disappeared, and only the more slowly diffusing water signal in brain tissue is visible.

randomly sampled using a uniform random number generator.

In the images shown in Fig. 1, the brain is surrounded by formalin in water. Formalin is an aqueous solution and the free diffusion of water in formalin is faster than in the brain tissue. One can tell that these images are sensitive to diffusion, because the signal from the water in formalin has disappears as the gradients increase.

THE DIFFUSION TENSOR MODEL

In isotropic diffusion, the solution of the diffusion equation is a Gaussian that gives the probability that a water molecule will diffuse a given distance. However, in anisotropic media, like baboon brain, diffusion is directionally dependent. The reason for this is simple, in brain nerve fibers have orientations, and diffusion along the fiber is different from diffusion across the fiber. In anisotropic media, the probability that a water molecule will diffuse a given distance is given by a symmetric three dimensional Gaussian distribution. The Gaussian is symmetric because the rate of diffusion from a to b is the same as the rate of diffusion from b to a; i.e., no flow. Consequently, the Gaussian has a total of 6 independent parameters. In the reference frame of the diffusion gradients, this Gaussian is related to the diffusion data by

$$d_i = A \exp\left\{-\kappa g_i \cdot \mathscr{T} \cdot g_i^\dagger\right\} + C + n_i \tag{1}$$

where the d_i are the intensities of one pixel from *all* 45 of the images. The amplitude A represents an arbitrary scale introduced by the spectrometer. The diffusion tensor, $\mathscr{T}$, is defined as

$$\mathscr{T} = \begin{pmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{xy} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{pmatrix}, \tag{2}$$

where D_{uv} is the diffusion coefficient along the uv direction. The ith gradient vector is represented symbolically by g_i. The constant C may be thought of as the component of the magnetic resonance signal that arises from highly constrained water molecules. The noise in the ith data value has been represented symbolically by n_i.

The constant κ, appearing in Eq. (1), is a conversion factor and is given by

$$\kappa = \gamma^2 \delta^2 (\Delta - \delta/3) \tag{3}$$

where γ is the magnetogyric ratio of of the nucleus of interest, δ is the duration of the diffusion encoding gradient and Δ is a delay between the onset of two diffusion encoding gradients. The two delay times are controlled by the experimenter when they setup the experiment. The magnetogyric ratio is a characteristic of the nuclide being observed, ^{1}H in this case.

In writing this model as a single pixel model we have made a number of simplifying assumptions. First, magnetic resonance images are taken in the time domain. The image is formed by taking a discrete Fourier transform. The discrete Fourier transform is a information preserving transform, so one can work either in the time domain or in the image domain and the Bayesian calculations are essentially unchanged. However, assuming that adjacent pixels do not interact is an approximation.

Magnetic resonance images are images of spin density as modified by relaxation decay of the signal. Spin densities are strictly positive real quantities having zero imaginary part. However, the discrete Fourier transform of the magnetic resonance data have nonzero positionally dependent phases. These phases vary linearly with position and they mix the real and imaginary parts of the discrete Fourier transform. The three phase parameters needed to unmix the real and imaginary parts of the discrete Fourier transform can in principle be estimated and the effects of these phases removed. The second simplifying assumption is that we can perform these phase calculations independent of the the Bayesian calculations presented in this paper. The Bayesian calculations needed to estimate the three phases are given in [1].

The images shown in Fig. 1 are the real part of these phased images, i.e., they are spin density maps. The imaginary parts of these images, not shown, have signal-to-noise ratios of less than one, indeed essentially zero, except where there are artifacts in the images, for example at sharp boundaries in the image. The third simplifying assumption is that there is no need to model the imaginary part of these images.

The diffusion tensor, $\mathscr{T}$, is a real symmetric matrix. Real symmetric matrices may be diagonalized using the eigenvalues and eigenvectors of $\mathscr{T}$. The eigenvalues of $\mathscr{T}$ will be designated as λ_1, λ_2 and λ_3. These eigenvalues are the magnitude of the diffusion along the three principle directions of the diffusion tensor. These principle directions are specified by the eigenvectors of the diffusion tensor. These eigenvectors form a unitary rotation matrix, R. Rotation matrices in three dimensions are characterized by three Euler angles, ϕ, θ, and ψ. This rotation matrix R may be written by a series of three rotations given by

$$R \equiv \begin{pmatrix} \cos\psi & \sin\psi & 0 \\ -\sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \cos\theta & 0 & -\sin\theta \\ 0 & 1 & 0 \\ \sin\theta & 0 & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\phi & \sin\phi & 0 \\ -\sin\phi & \cos\phi & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad (4)$$

where the right most matrix is a rotation about z through an angle ϕ, followed by a rotation about y through an angle θ (center matrix). Finally, the third rotation, left most matrix, is a rotation about z through an angle ψ. When multiplied out the rows of this matrix are the eigenvectors needed to diagonalized the diffusion tensor.

TABLE 1. The Diffusion Tensor Models

Model Indicator	Model Name	$A=0$	$C=0$	Eigenvectors	Angles
1	No Signal	Yes	Yes	N/A	None
2	Constant	Yes	No	N/A	None
3	Isotropic	No	Yes	$\lambda_1=\lambda_2=\lambda_3$	None
4	Isotropic+Const	No	No	$\lambda_1=\lambda_2=\lambda_3$	None
5	Pancake	No	Yes	$\lambda_1<\lambda_2=\lambda_3$	$\psi=0$
6	Pancake+Const	No	No	$\lambda_1<\lambda_2=\lambda_3$	$\psi=0$
7	Football	No	Yes	$\lambda_1>\lambda_2=\lambda_3$	$\psi=0$
8	Football+Const	No	No	$\lambda_1>\lambda_2=\lambda_3$	$\psi=0$
9	Full Tensor	No	Yes	$\lambda_1>\lambda_2>\lambda_3$	$\psi\neq 0$
10	Full Tensor+Const	No	No	$\lambda_1>\lambda_2>\lambda_3$	$\psi\neq 0$

Using the eigenvalues and the rotation matrix, R, the diffusion tensor model, Eq. (1), may be transformed into

$$d_i = A\exp\left\{-\kappa g_i\cdot RVR^{\dagger}\cdot g_i^{\dagger}\right\}+C+n_i. \tag{5}$$

The matrix V is a diagonal matrix and is given by

$$V\equiv\begin{pmatrix}\lambda_1 & 0 & 0\\ 0 & \lambda_2 & 0\\ 0 & 0 & \lambda_3\end{pmatrix}. \tag{6}$$

For a full diffusion tensor model, the parameters of interest would be the three eigenvalues, the Euler angles, the amplitude and the constant (if present).

In the problem we are addressing, the models we wish to test are specified by imposing symmetries on the diffusion tensor. For example, if the diffusion is isotropic then $\lambda_1=\lambda_2=\lambda_3$ and the diffusion spherical. In spherical diffusion all of the Euler angles are zero and the rotation matrix is one. Similarly, if the diffusion is prolate (football shaped), then $\lambda_2=\lambda_3$, the diffusion is symmetric about its long axis and the angle ψ is zero. The full list of models is given in Table 1.

For more on the diffusion tensor model see [2, 3, 4]. There you will find an extensive explanation of the diffusion tensor model.

THE BAYESIAN CALCULATIONS

The problem is to compute the posterior probability for the model indicator given the diffusion tensor data using Bayesian probability theory [5, 6]. There are 10 models and so 10 calculations that must be done. However, all of the calculations are essentially identical and we give only the calculation for the full diffusion tensor model plus a constant. The other calculations may be obtained by imposing the appropriate symmetries and removing the priors for any parameters that do not occur. The posterior probability for the model indicator, u, is represented symbolically by $P(u|DI)$. This posterior

probability is computed by application of Bayes' theorem [7], one obtains

$$P(u|DI) = \frac{P(u|I)P(D|uI)}{P(D|I)} \tag{7}$$

where $P(u|I)$ is the prior probability for the model indicator, which we assigned using a uniform prior probability. $P(D|uI)$ is the direct probability for the data given the model indicator, and $P(D|I)$ is a normalization constant and is given by

$$P(D|I) = \sum_{u=1}^{10} P(uD|I) = \sum_{u=1}^{10} P(u|I)P(D|uI). \tag{8}$$

If we normalize $P(u|DI)$ at the end of the calculation, then the posterior probability for the model indicator is proportional to the direct probability for the data given the model indicator:

$$P(u|DI) \propto P(D|uI). \tag{9}$$

This direct probability is a marginal probability and is given by

$$\begin{aligned} P(D|uI) &= \int d\Omega P(\Omega D|uI) \\ &= \int d\Omega P(\Omega|uI)P(D|\Omega uI) \end{aligned} \tag{10}$$

where we are using Ω to stand for all of the parameters appearing in the model. For the full diffusion tensor plus a constant model, $\Omega \equiv \{A, C, \theta, \phi, \psi, \lambda_1, \lambda_2, \lambda_3, \sigma\}$. We have added one additional parameter, σ, to this list. This additional parameter represents what is known about the noise. Factoring the prior probability for the parameters into independent prior probabilities for each parameter one obtains:

$$\begin{aligned} P(D|uI) &= \int d\Omega P(A|I)P(C|I)P(\theta|I)P(\phi|I)P(\psi|I)P(\sigma|I) \\ &\times \quad P(\lambda_1|I)P(\lambda_2|I)P(\lambda_3|I)P(D|\Omega uI). \end{aligned} \tag{11}$$

The prior probability for σ, $P(\sigma|I)$, was assigned using a Jeffreys' prior probability and σ was removed by marginalization. Strictly speaking, to use a Jeffreys' prior probability one must bound and normalize the Jeffreys' prior and then at the end of the calculation allow these bounds to go off to infinity as a limit. However, in this calculation the parameter σ appears in each model in exactly the same way and so any normalization constant associated with bounding and normalizing the Jeffreys' prior cancels when these model probabilities are normalized. Additionally, the Student's t-distribution that results from removing σ is so strongly convergent that the use of a Jeffreys' prior probability is harmless in this problem.

All other prior probabilities were assigned using fully normalized prior probabilities. The prior probabilities for the three angles were assigned using uniform prior probabilities. The prior probabilities for the amplitude, the constant and the three eigenvalues were assigned using normalized Gaussian prior probabilities. If X represents one of these

parameters, then

$$P(X|H_X L_X) = \begin{cases} \left(2\pi \mathrm{Sd}_X^2\right)^{-\frac{1}{2}} \exp\left\{-\dfrac{(\mathrm{Mean}_X - X)^2}{2\mathrm{Sd}_X^2}\right\} & \text{If } L_X \leq X \leq H_X \\ 0 & \text{otherwise} \end{cases} \tag{12}$$

where L_X and H_X are the low and high parameter values. The 'Mean' value of the Gaussian prior probability was set to the center of the low-high interval:

$$\mathrm{Mean}_X = (L_X + H_X)/2. \tag{13}$$

The standard deviation of this Gaussian was set so that the entire interval, low to high, represents a 3 standard deviation interval:

$$\mathrm{Sd}_X = (H_X - L_X)/3. \tag{14}$$

The model equation is symmetric under exchange of labels on the eigenvalues. This symmetry manifests itself in the posterior probability. If there is a peak in the posterior probability at $\lambda_1 = a$ and $\lambda_2 = b$ then there is also a peak at $\lambda_1 = b$ and $\lambda_2 = a$. Consequently, the prior probabilities for the eigenvalues were assigned using Eq. (12) subject to the additional condition $\lambda_1 > \lambda_2 > \lambda_3$; which breaks this symmetry and leaves a single global maximum in the posterior probability. This condition is equivalent to defining what we mean by eigenvalue one: we mean the largest eigenvalue.

The direct probability for the data given the parameters, $P(D|\Omega uI)$, was assigned using a Gaussian prior probability for the noise:

$$P(D|\Omega uI) = (2\pi\sigma^2)^{-\frac{N}{2}} \exp\left\{-\frac{Q}{2\sigma^2}\right\} \tag{15}$$

where N is the total number of data values (images) and Q is given by

$$Q = \sum_{i=1}^{N} \left(d_i - A\exp\left\{-c g_i^{\dagger} \cdot RVR^{\dagger} \cdot g_i\right\} - C\right)^2. \tag{16}$$

Using Eq. (15), keeping the prior probabilities in symbolic form, and evaluating the integral over the standard deviation of the noise, the posterior probability for the model indicator is given by

$$\begin{aligned} P(u|DI) &\propto P(D|uI) \\ &\propto \int dA dC d\theta d\phi d\psi d\lambda_1 d\lambda_2 d\lambda_3 P(A|I)P(C|I)P(\theta|I)P(\phi|I)P(\psi|I) \\ &\times \quad P(\lambda_1|I)P(\lambda_2|I)P(\lambda_3|I)\left[\frac{Q}{2}\right]^{-\frac{N}{2}}. \end{aligned} \tag{17}$$

This equation is the solution to the model selection calculation for the full diffusion tensor plus a constant model. As noted, to obtain the posterior probability for the other

models all one needs to do is to apply the appropriate symmetries to this model. For example, the posterior probability for the isotropic diffusion tensor, $u = 3$, has equal eigenvalues, no angles and no constant. Consequently, the model equation is given by:

$$d_i = A \exp\left\{-\kappa g \cdot g^{\dagger} \lambda_1\right\} + n_i \tag{18}$$

where $g \cdot g^{\dagger}$ is the total squared length of the gradients, and only a single eigenvalue, λ_1, is present. Additionally, for the isotropic case, the rotation matrix, R, is the identity matrix and has been multiplied out. Using this model and imposing the appropriate symmetries on Eq. (17), the posterior probability for the isotropic diffusion tensor is given by

$$P(u = 3|DI) \propto \int dA d\lambda_1 P(A|I) P(\lambda_1|I) \left[\frac{Q}{2}\right]^{-\frac{N}{2}} \tag{19}$$

where Q is now given by

$$Q \equiv \sum_{i=1}^{N} \left(d_i - A \exp\left\{-\kappa g \cdot g^{\dagger} \lambda_1\right\}\right)^2. \tag{20}$$

By imposing the appropriate symmetries on Eq. (17), the posterior probabilities for all of the other models may be easily derived.

DISCUSSION

Implementing this calculation can be formidable because evaluation of the integrals in Eq. (17) is highly nontrivial. These integrals vary in complexity from no integral, the 'No Signal' model, to as complicated as an 8 dimensional integral, the full diffusion tensor plus an constant model. Nonetheless, this is the calculation that we implement using a Metropolis-Hastings Markov chain Monte Carlo simulation with simulated annealing.

In the program that implements the calculation we try to keep the Markov chain Monte Carlo simulations at a stationary point. By a stationary point we mean that for a given value of the annealing parameter the expected value of the parameters and the logarithm of the likelihood are stationary. Stationary in the sense that we can run the Markov chain over many cycles taking as many samples as we please and neither the mean value of the parameters nor the mean value of the logarithm of the likelihood change.

The model indicator is treated just like any other parameter. The model indicator is varied using a Gaussian proposal that allows the model indicator to move as much as 3 at one standard deviation and by as much as 6 at two standard deviations. Consequently, the program can quickly explore the entire model space. However, unlike most parameters when the model indicator changes, the number of parameters may change. Additionally, because of the exponential nature of the diffusion tensor, the region of parameter space corresponding to the stationary point changes when the model indicator changes. One must have a scheme for proposing diffusion tensor parameters that ensures the simulation is at or near a stationary point for a given value of the annealing parameter. Cloning diffusion tensor parameters, i.e., finding a diffusion tensor of the proposed

type and copying its parameters, is not a good option because for high values of the annealing parameter most proposed model indicators are not present in the numerical simulations. We solved this proposal problem by initializing all ten model from the priors. When the program proposes a new model indicator it switched to the indicated set of parameters. These parameters are then simulated until they reach equilibrium. The simulation is either accepted or rejected using the standard Metropolis-Hastings criteria. For low values of the annealing parameter, changes in model indicator are readily accepted. Consequently, the parameters associated with a given model indicator are never very far from equilibrium. So, simulating the proposed model until the parameters reach equilibrium is not as time consuming as proposing the parameters from the prior. See [8, 9, 10] for more on Markov chain Monte Carlo and see [11, 12] for more on how to use simulated annealing to perform model selection calculations.

A typical NMR image might have 32,786 pixels per slice, so the numerical calculations for even a single slice are formidable. As a result, the calculations were implemented in parallel using the so called shared memory model. Parallelization occurs at the pixel level, so multiple pixels are processed simultaneously. We run the calculations on a 64 node Silicon Graphics Altix 3000 supercomputer running the Intel Itanium 2 processors. A typical slice requires roughly 2 hours elapse time to analyze using 32 processors. Figure 2 is one example of the types of outputs obtained using the data shown in Fig. 1. This particular output is the expected model indicator and we will have more to say about this figure shortly.

The program was tested on simulated and real data. In the tests using simulated data, we generated data having known model indicators and parameters with signal-to-noise ratio of about 30, a signal-to-noise ratio typical of real data. Then using this data we ran the analysis to see how well the model selection worked. In these data the identification is near 100% with errors occurring only when the generated data looks like one of the submodes. For example, if the generated data was full diffusion tensor having λ_2 nearly equal to λ_3 then the program understandably identifies the data as a football diffusion tensor rather than a full diffusion tensor.

The program that implements the Bayesian calculation is a Metropolis-Hastings Markov chain Monte Carlo simulation. Consequently, it has samples from each marginal posterior probability for each parameter appearing in each high probability model. However, the program does not output these samples directly; rather images of various averages of these samples are output. For example, Fig. 2 is the expected model indicator. The expected model indicator is computed as

$$\langle u \rangle = \sum_{u=1}^{10} uP(u|DI) \tag{21}$$

where the posterior probability for the model indication, $P(u|DI)$, is computed from the Markov chain Monte Carlo samples.

If you examine Fig. 2 you will discover that outside the sample container, the mean model is the 'No Signal' model, as it should be. Inside the sample container, in the space occupied by formalin, the mean model is the 'Isotropic' diffusion model. Again this is the how one would hope the expected model would behave. Inside the brain tissue, the expected model is more complicated. However, we can say that the expected model

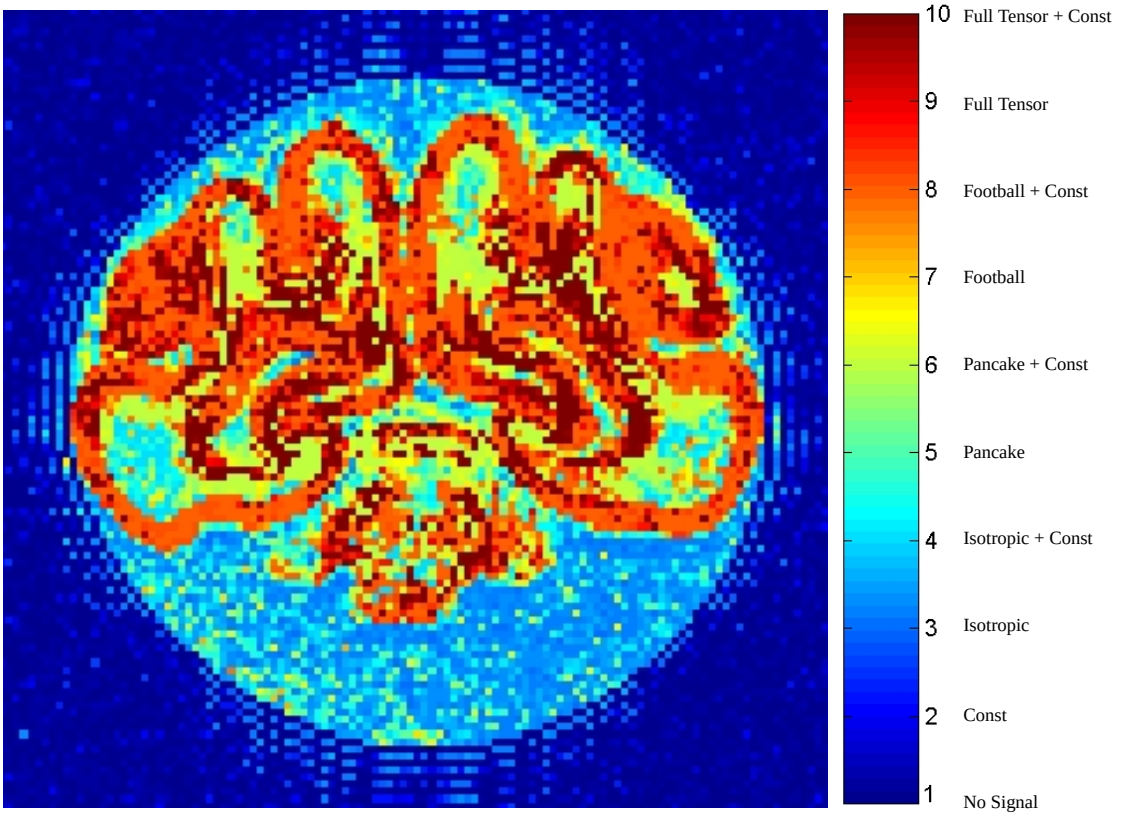

FIGURE 2. The Expected model value behaves as expected: outside the sample container the 'No Signal' model is selected, while in the container in the formalin the 'isotropic' model is selected. Finally inside the brain tissue the model selected are more complex and follow the anatomical structures of the brain.

always contains a constant, Fig. 3, and in the cortex the expected model is a football plus a constant model.

Figure 3 is an image of the fraction of the total signal contained in the constant. The expected fractional constant is computed in a way analogous to Eq. (21). The mean fractional constant is computed for each of the 10 models and these mean fractional constants are then weighted by the posterior probability for each model. If f_u is the mean fractional constant computed from the Markov chain Monte Carlo samples for the uth model, then the expected fractional constant is given by

$$\langle f \rangle = \sum_{u=1}^{10} f_u P(u|DI). \tag{22}$$

The fractional constant is defined as the the ratio of the constant to the total signal intensity. If a model does not contain a constant, for example the isotropic model, then the expected fractional constant is defined to be zero. The fractional constant image looks as if we have cropped the image. However, this is not the case. The image shown

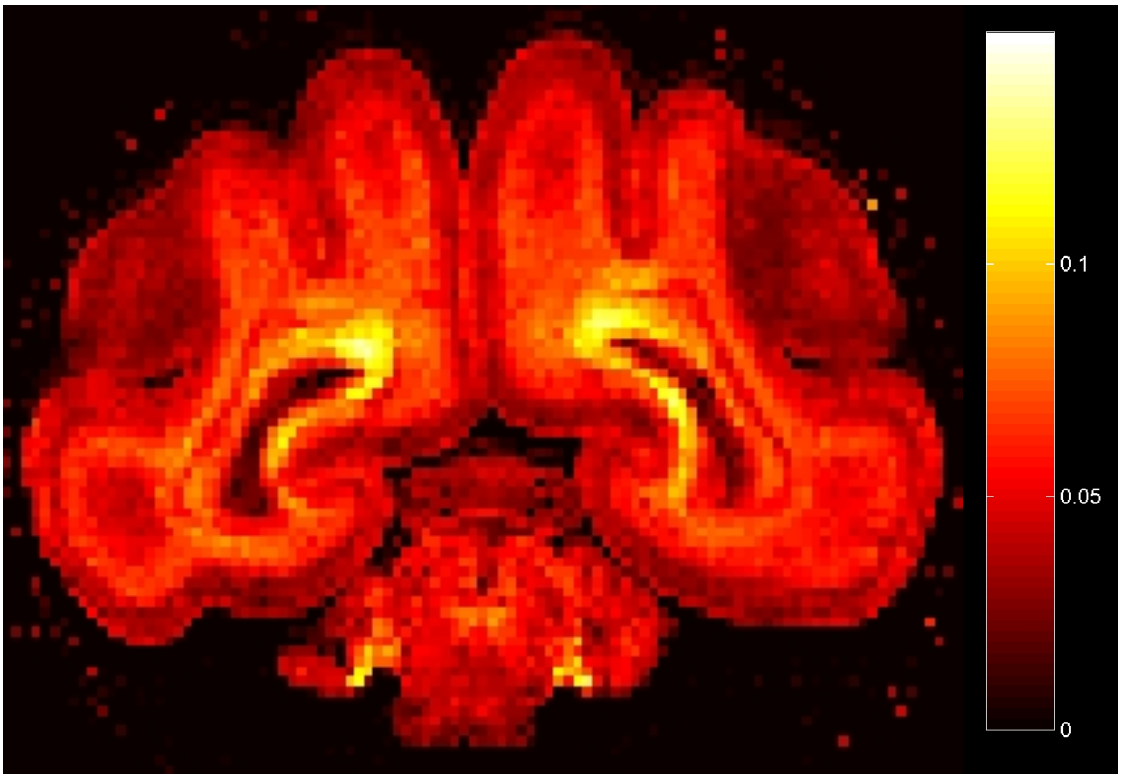

FIGURE 3. The expected fractional constant is the constant amplitude divided by the total signal amplitude. It is an expected value in that it is a weighted average computed for each model weighted by the posterior probability for that model. Models that do not contain a constant, by definition, have a constant equal to zero.

in Fig. 3 is exactly what comes out of the calculation. In regions where there is 'No Signal' there is no constant and, so, no fractional constant. Similarly, in regions where the diffusion is isotropic, for example in the formalin, there is no constant and so no fractional constant. The only region where a diffusion tensor plus a constant model is selected is in the brain tissue, and as this image illustrates, in brain the constant is always selected.

In addition to computing the expected fractional constant, the program also outputs the expected average diffusion coefficient, Fig. 4, and the expected fractional anisotropy, Fig. 5. In the brain the expected value of the average diffusion coefficient does not have a lot of structure. However, had we not included the constant this would not have been the case. It is always important to have the correct model, and this is especially important when dealing with exponentials. If an exponential model is not correct, the parameter estimates one obtains from that model will not reflect the "true" decay of the signal, [13]. In this analysis if the constant had not been included the expected value of the diffusion coefficient would reflect the constant offset; for a fixed amplitude, increasing the constant will decrease the expected diffusion coefficient, similarly decreasing the constant causes the diffusion coefficient to rise, i.e., the signal must decay faster.

The expected fractional anisotropy, Fig. 5, is the standard deviation of the three

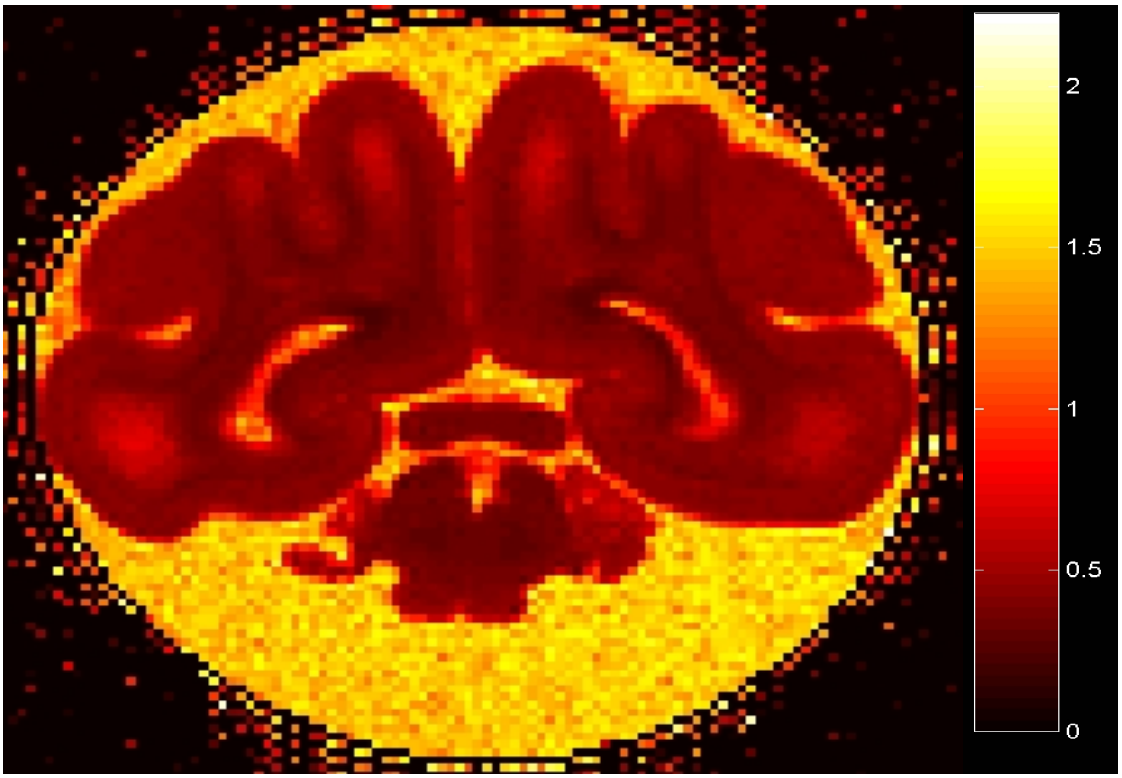

FIGURE 4. The expected average diffusion coefficient is also output from the program. It is a weighted average computed from the average value of the eigenvalue weighted by the posterior probability for that model. By definition the 'No Signal' has zero diffusion coefficient.

eigenvalues, the λ's, divided by the average eigenvalue. By definition the 'No Signal' and all the 'Isotropic' models have zero fractional anisotropy. As with all of the other outputs discussed, the mean fractional anisotropy is computed for each of the 10 models, and then the weighted average is formed by multiplying each mean fractional anisotropy by the posterior probability for the model. As discussed with the fractional constant, the fractional anisotropy only has nonzero values when an anisotropic diffusion tensor model is selected. In this case that means inside the brain and then only in regions where the model is not isotropic. In this brain the fractional anisotropy is large only in the cortex (at the outer surface of the brain). In the underlying developing white mater, the fractional anisotropy is almost zero.

SUMMARY AND CONCLUSIONS

Bayesian probability theory has been used to compute the posterior probability for the model indicators given the 10 models shown in Table 1. The calculation is implemented using a Metropolis-Hastings Markov chain Monte Carlo simulation with simulated annealing. In this calculation the model indicator is treated as a discrete parameter and sampled in a way exactly analogous to any other parameter in a Markov chain.

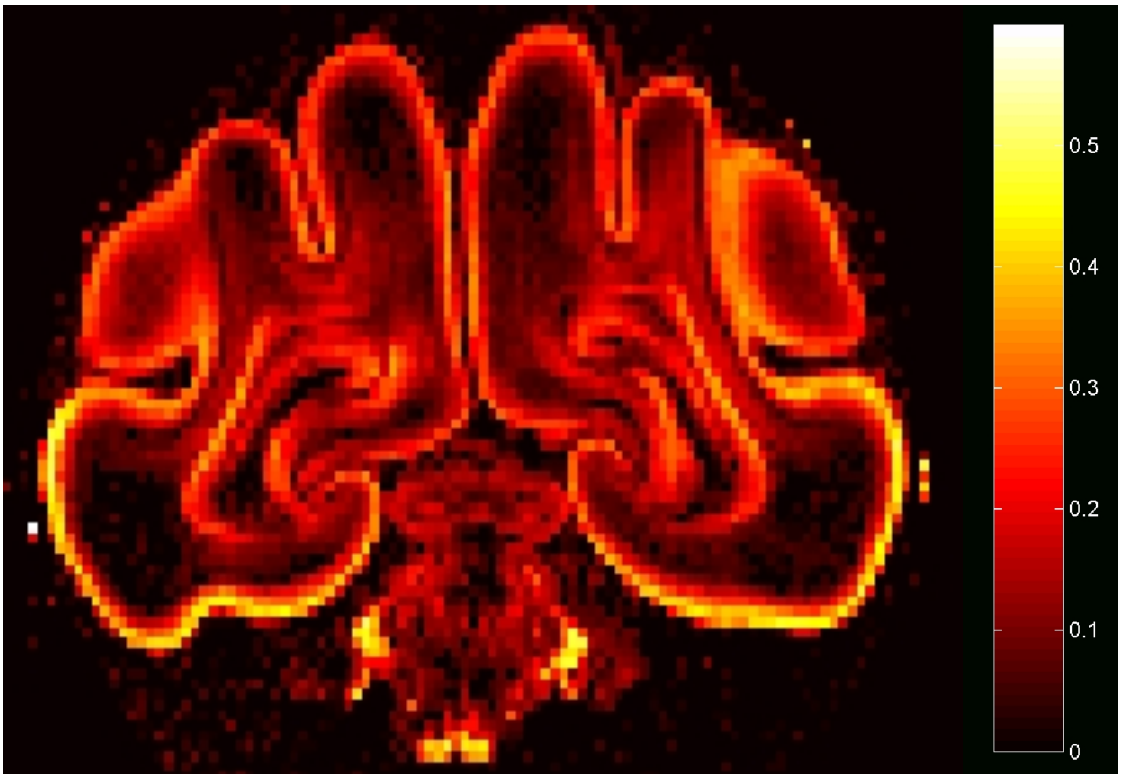

FIGURE 5. The expected fractional anisotropy is the standard deviation of the three eigenvalues, the λ's, divided by the average eigenvalue. It is a weighted average computed from the eigenvalues from each model weighted by the posterior probability for that model. By definition the 'No Signal' and all the 'Isotropic' models have zero fractional anisotropy.

New model indicators are proposed. The parameters are simulated until they reach equilibrium at a given value of the annealing parameter and the model indicator is either accepted or rejected according to the Metropolis-Hastings algorithm. As noted in the introduction, we are now in the process of applying these calculations to diffusion tensor images of fixed baboon brains as a function of gestational age. This work, to be reported elsewhere, is giving us new insights into the diffusion of water molecules and how the maturation of the brain affects diffusion.

ACKNOWLEDGMENTS

This work was supported by a contract with Varian NMR Systems, Palo Alto, CA; by the Small Animal Imaging Resources Program (SAIRP) of the National Cancer Institute, grant R24CA83060; and by the National Institute of Neurological Disorders and Stroke grants NS35912 and NS41519.

REFERENCES

1. G. Larry Bretthorst, "Automatic Phasing of NMR Images Using Bayesian Probability Theory," *J. Magn. Reson.*, in preparation.
2. Basser, P. J., J. Mattiello, D. Le Bihan (1994), "MR diffusion tensor spectroscopy and imaging," *Biophys. J.*, **66,** 259-267.
3. Basser, P. J., J. Mattiello, D. Le Bihan (1997), "Estimation of the effective self-diffusion tensor from the NMR spin echo," *J. Magn. Reson. Ser. B*, **103,** 247-254.
4. Thomas E. Conturo, Robert C. McKinstry, Erbil Akbudak and Bruce H. Robinson (1996), "Encoding of Anisotropic Diffusion with Tetrahedral Gradients: A General Mathematical Diffusion Formalism and Experimental Results, *Magnetic Resonance in Medicine*, **35**, pp 399-412.
5. E. T. Jaynes (2003), "Probability Theory—The Logic of Science," G. L. Bretthorst, Ed., Cambridge University Press, Cambridge, UK.
6. Bretthorst, G. Larry (1996), "An Introduction To Model Selection Using Bayesian Probability Theory," in *Maximum Entropy and Bayesian Methods*, G. R. Heidbreder, ed., pp. 1–42, Kluwer Academic Publishers, the Netherlands.
7. Bayes, Rev. T. (1763), "An Essay Toward Solving a Problem in the Doctrine of Chances," *Philos. Trans. R. Soc. London* **53,** pp. 370-418; reprinted in *Biometrika* **45,** pp. 293-315 (1958), and *Facsimiles of Two Papers by Bayes*, with commentary by W. Edwards Deming, New York, Hafner, 1963.
8. N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, E. Teller (1953) "Equations of state calculations by fast computing machines," *J. Chem. Phys.* **21**, pp. 1087-1091.
9. W. R. Gilks, S. Richardson and D. J. Spiegelhalter (1996) "Markov Chain Monte Carlo in Practice," Chapman & Hall, London.
10. R. M. Neal (1993) "Probabilistic inference using Markov chain Monte Carlo methods," technical report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.
11. J. Skilling (1998) "Probabilistic data analysis: and introductory guide," *J. of Microscopy*, **190**, Pts 1/2, pp.28̃-36.
12. Paul M. Goggans, and Ying Chi (2004), "Using Thermodynamic Integration to Calculate the Posterior Probability in Bayesian Model Selection Problems," in *Bayesian Inference and Methods in Science and Engineering*, Joshua Rychert, Gary Erickson and C. Ray Smith *(eds.)*, American Institute of Physics, USA, (in press).
13. G. Larry Bretthorst, W. C. Hutton, J. R. Garbow, W. C. Hutton, J. J. H. Ackerman (2004) "Data Analysis In NMR Using Bayesian Probability Theory I. Parameter Estimation," *J. Magn. Reson.*, in press.