NeuroImage 45 (2009) 1183–1189

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Technical Note

A general method of Bayesian estimation for parametric imaging of the brain

Nathaniel M. Alpert *, Fang Yuan

Division of Nuclear Medicine and Molecular Imaging, Department of Radiology, Massachusetts General Hospital, 50 Fruit Street, Boston, MA 02114, USA

ARTICLE INFO

Article history:
Received 3 July 2008
Revised 15 December 2008
Accepted 31 December 2008
Available online 14 January 2009

ABSTRACT

We report a general method of Bayesian estimation that uses prior measurements to improve the signal-to-noise ratio of parametric images computed from dynamic PET scanning. In our method, the ordinary weighted least squares cost function is augmented by a penalty term to yield $\Phi(\boldsymbol{K},S) = \min_{\boldsymbol{K}}\{(\boldsymbol{C}-f(\boldsymbol{K}))^T\Omega_{\boldsymbol{C}}^{-1}(\boldsymbol{C}-f(\boldsymbol{K})) + S\Phi(\boldsymbol{K},S=0)(\boldsymbol{K}-\hat{\boldsymbol{K}})^T W_{\boldsymbol{K}}^{-1}(\boldsymbol{K}-\hat{\boldsymbol{K}})\}$, where C is a PET concentration history and $\boldsymbol{\Omega_C}$ is its variance, f is the model of the concentration history, $\boldsymbol{K}=[k_1,k_2,\ldots,k_m]^T$ is the parameter vector, $\hat{\boldsymbol{K}}$ is the vector of population means for the model parameters, $\boldsymbol{\Omega_K}$ is its covariance, $\Phi_K(\boldsymbol{K},S=0)$ is the conventional weighted sum of squares. $S>0$ is chosen to control the balance between the prior and new data. Data from a prior population of subjects are analyzed with standard methods to provide maps of the mean parameter values and their variances. As an example of this approach we used the dynamic image data of 10 normal subjects who had previously been studied with ^{11}C-raclopride to estimate the prior distribution. The dynamic data were transformed to stereotactic coordinates and analyzed by standard methods. The resulting parametric maps were used to compute the voxel-wise sample statistics. Then the cohort of priors was analyzed as a function of S, using nonlinear least squares estimation and the cost function shown above. As S is increased the standard error in estimating BP in single subjects was substantially reduced allowing measurement in BP in thalamus, cortex, brain stem, etc. Additional studies demonstrate that a range of S values exist for which the bias is not excessive, even when parameter values differ markedly from the sample mean. This method can be used with any kinetic model so long as it is possible to compute a map of a priori mean parameters and their variances.

Introduction

There are a number of standard dynamic measurements that can be made with PET, including endpoints such as cerebral blood flow, glucose metabolism, and the binding indices of D2, DAT and other ligands. When analyzing individual studies of animal or human subjects, low signal-to-noise ratio (SNR) is an impediment to detailed analysis of the three-dimensional distribution of the endpoints of interest. Many, if not most, analyses rely on a kinetic model and least squares parameter estimation. When a kinetic model is fit to noisy data a large number of solutions may have similar sums of squares, including some fits with unphysiological parameter values. As a consequence, estimates of the model parameter have large uncertainties. In fact, this is a common problem with parametric imaging because the kinetic curves generated from single voxels are often quite noisy. In PET studies with short lived tracers, such as ^{11}C- and ^{18}F-labeled compounds, the SNR of the concentration values usually peaks early in the study and then tends to fall with time. Statistical noise in the PET concentration data are propagated into the parametric images. A number of techniques have been developed to increase the effective SNR of the dynamic voxel-level data. Many of these techniques use some type of spatial and/or temporal smoothing to increase the SNR of the voxel level kinetic curves, followed by weighted nonlinear least squares parameter estimation. Wavelet denoising has recently been shown to be effective in increasing SNR of the PET data without substantial loss of resolution (Millet et al., 2000; Cselenyi et al., 2002; Turkheimer et al., 2003; Alpert et al., 2006; Shidahara et al., 2007) Clustering of voxel histories with similar shape has also been used to increase the SNR of the kinetic curves (Kimura et al., 1996; 1999).

This paper focuses on a method to reduce the propagation of noise in the estimation of parametric images. We recognize that considerable work has already been done and substantial progress has been made but most of the work has involved the method of least squares. Adding Bayesian estimation methods offers the possibility of significant improvements.

Least squares estimation has a long history in science and engineering and many good reviews are available (e.g. Bard, 1974; Bevington and Robinson, 2003). The estimation problem always involves a tradeoff between the accuracy and precision of the parameter estimates. Muzic and Christian (2006) have recently compared a number of estimation methodologies and found that when analyzing noisy kinetic data, the frequently used weighted least squares method did not perform as well as iteratively reweighted least squares or extended least squares. Until recent years, research in

* Corresponding author. Fax: +1 617 726 6165.
E-mail address: alpert@pet.mgh.harvard.edu (N.M. Alpert).

1053-8119/$ – see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.neuroimage.2008.12.064

kinetic modeling and parametric imaging focused on computing minimum variance, unbiased estimators. But investigators have shown that by accepting some bias, it is possible to achieve much lower parameter variance. An attractive heuristic is that due to the finite resolution of the PET scanner, the parametric images values should not vary wildly from voxel to voxel. This suggests that methods falling under the general classification of penalized least squares may perform better than the conventional method. Zhou et al. (2001, 2002, 2003) and others (O'Sullivan and Saha, 1999) have successfully used biased estimation methods, such as ridge regression, to "regularize" parametric images.

There is another heuristic approach that seems appealing — the use of prior experience or information to adjust the estimates. The most famous example of this approach is Bayes' rule

$$P(\boldsymbol{C})P(H|\boldsymbol{C}) = P(\boldsymbol{C}|H)P(H) \tag{1}$$

Eq. (1) is a general symbolic formula but for the purpose of this paper we can identify the symbols as follows: $P(C|H)$ is the conditional probability of the PET concentration data, C, given the hypothesis, H. $P(C)$ is the a priori probability of the data. $P(H)$ is the a priori probability of the hypothesis. In our case, the hypothesis is expressed as a particular kinetic model, f, with parameter vector $\boldsymbol{K}$ which predicts the experiment experimental data as $f(\boldsymbol{K})$. In the following development $P(\boldsymbol{K}|C)$ denotes the conditional probability of the hypothesis H given the data.

Bayes' theorem can be used to extend the method of weighted least squares. Earlier applications to tracer kinetics and dynamic MR include those by Chen et al. (1988) and Yu et al. (1995) but to our knowledge it has not been applied to parametric imaging. A brief review of the theory of Bayesian regression is followed by the development of modifications needed for parametric imaging. As an example of Bayesian regression we measure raclopride binding potential in striatal and extrastriatal brain regions from parametric images but we emphasize that the technique is completely general with respect to kinetic model.

Methods

To describe the theory, we employ a notation where $\boldsymbol{C}$ is a vector representing the concentration history in a representative voxel from a four-dimensional PET image data set. In effect, $\boldsymbol{C}$ represents a concentration history from a single voxel, with m elements, namely $\boldsymbol{C}=[c_1,c_2,\ldots,c_m]^T$. The kinetic model, f, has r parameters denoted by $\boldsymbol{K}=[k_1,k_2,\ldots,k_r]^T$, with $r<m$.

Consider the likelihood of measuring the concentration history in a representative voxel of the image array. The concentration data may be spatially correlated due to the process of image reconstruction, but due to the nature of radioactive decay, the temporal data for any voxel, measured in separate time epochs (i.e. frames), are statistically independent. For this work we assume each datum is normally distributed with expected values $\hat{\boldsymbol{C}} = \left[\hat{c}_1,\hat{c}_2,\ldots,\hat{c}_m\right]^T$ and variances $\sum = [\sigma_1^2,\sigma_2^2,\ldots,\sigma_m^2]^T$. The statistical model is written as

$$\boldsymbol{C} = f(\boldsymbol{K}) + \varepsilon \tag{2}$$

where ε is an $m\times 1$ row vector of normally distributed random noise with mean zero and variance $\sum$. The expected values of the concentrations are generally unknown but they are modeled as known functions of the parameters, $f(\boldsymbol{K})$. Now, suppose that in addition to the experimental data, we have a cohort of prior data, for example an ensemble of parametric images. As we explain later, the ensemble can be used to compute the voxel-by-voxel sample mean and covariance of the parametric images, yielding an estimate of the population mean and covariance. Assume that the prior distribution of parameters can be described by a multivariate Gaussian distribution with the mean for a representative voxel given by $\hat{\boldsymbol{K}} = \left[\hat{k}_1,\hat{k}_2,\ldots,\hat{k}_r\right]^T$ and its population covariance matrix given by $\Omega_{\boldsymbol{K}_j}$, an r by r array. The theory of maximum likelihood can then be used to obtain the "best" estimate of the parameters by maximizing $P(\boldsymbol{K}|\boldsymbol{C})$. Since the concentration history is assumed to be normally distributed and statistically independent,

$$P(\boldsymbol{C}|\boldsymbol{K}) = \frac{\sqrt{\det(\Omega_{\boldsymbol{C}}^{-1})}}{(2\pi)^{m/2}}\exp\left(-\frac{1}{2}(\boldsymbol{C}-f(\boldsymbol{K}))^T W_{\boldsymbol{C}}^{-1}(\boldsymbol{C}-f(\boldsymbol{K})\right), \tag{3}$$

where $\Omega_{\boldsymbol{C}} = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_m^2 \end{bmatrix}$ and

$$P(\boldsymbol{K}) = \sqrt{\frac{\det(\Omega_{\boldsymbol{K}}^{-1})}{(2\pi)^{r/2}}}\exp\left(-\frac{1}{2}\left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)^T W_{\boldsymbol{K}}^{-1}\left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)\right), \tag{4}$$

Substituting according to Eq. (1) and dividing by $P(\boldsymbol{C})$ yields the likelihood function

$$P(\boldsymbol{K}) = \frac{1}{\sqrt{\det(\Omega_{\boldsymbol{C}})}P(\boldsymbol{C})(2\pi)^{(r+m)/2}} \times \exp\left(-\frac{1}{2}(\boldsymbol{C}-f(\boldsymbol{K}))^T + \frac{1}{2}\left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)^T W_{\boldsymbol{K}}^{-1}\left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)\right) \tag{5}$$

Then, the Bayesian regression estimate of $\boldsymbol{K}$ is obtained by minimizing the log-likelihood

$$\Phi(\boldsymbol{K}) = \min_{\boldsymbol{K}}\left\{(\boldsymbol{C}-f(\boldsymbol{K}))^T\Omega_{\boldsymbol{C}}^{-1}(\boldsymbol{C}-f(\boldsymbol{K})) + \left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)^T\Omega_{\boldsymbol{K}}^{-1}\left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)\right\} \tag{6}$$

where, for simplicity, we have omitted factors independent of $\boldsymbol{K}$, as they do not affect the result. We see that Eq. (6) is made up of two terms, call them Ψ and β : The first term, Ψ, is the usual cost function for weighted least squares. The second term, β, is the "Bayes'" term, due to the inclusion of prior data.

There are two practical problems to solve before the theory of Bayesian Regression can be applied to PET and parametric images: 1. Additional approximations are required to make the application of Eq. (6) tractable. 2. The values of $\hat{\boldsymbol{K}}$ and $\boldsymbol{\Omega_K}$ must be determined for each voxel in the parametric image space.

The reason for additional approximations and development arise because the least squares term Ψ usually includes an arbitrary scale factor. This is due to the fact that $\boldsymbol{\Omega_C}$ is difficult to compute, a point we discuss in more detail later in this paper. Thus, the relationship between Ψ and β is not easily calculated. To overcome this problem we propose an approximate method by defining a scaling parameter for each voxel in the parametric image space. In the work that follows we use the approximation formula

$$\Phi(\boldsymbol{K},S) = \min_{\boldsymbol{K}}\left\{(\boldsymbol{C}-f(\boldsymbol{K}))^T\Omega_{\boldsymbol{C}}^{-1}(\boldsymbol{C}-f(\boldsymbol{K})) + S\Phi(\boldsymbol{K},S=0)\left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)^T\Omega_{\boldsymbol{K}}^{-1}\left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)\right\} \tag{7}$$

where S is a constant, independent of voxel location, chosen empirically by a process discussed later; whereas, $\Phi(\boldsymbol{K},S=0)$ is essentially the residual sum of squares after fitting the model by conventional weighted nonlinear least squares. $\Phi(\boldsymbol{K},S=0)$ is determined at each voxel by weighted least squares fitting; accordingly, the product, $S\Phi(\boldsymbol{K},S=0)$, is a scaling factor determining the (approximate) relationship between the Ψ and β terms in Eq. (6).

We included the effect of prior data in Bayesian regression by use of stereotactic sample statistics. The basic procedure requires a standard

kinetic model, with parameter estimation effected by least squares, and a representative cohort of subjects. In principle, the cohort should be large enough to support a reliable estimate of the sample distributions of the parameters. We assumed the parameters were normally distributed, characterized by their sample means and variances. Each subject in the cohort is analyzed with the standard kinetic method to form parametric images, one for each parameter in the model. The parametric images are spatially normalized to a standard size and spatial orientation, using the MNI stereotactic coordinate system (Collins, et al., 1994; Evans et al., 1997). The voxelwise sample mean and standard deviation of each parametric image type is formed and stored for later use. Note that we have employed the sample standard deviation, rather than the complete covariance matrix. This is an expedient simplification but not a fundamental aspect of our method.

To illustrate the use of prior data we provide a demonstration of Bayesian regression applied to parametric imaging of raclopride. A cohort of raclopride measurements were obtained from 11 young normal subjects participating in prior research studies. Ten subjects were used to construct the a priori data and one data set was set aside to further explore the use of Bayesian regression. The details of subject selection and data acquisition are given in Badgaiyan et al. (2007). Briefly, subjects in this study were young healthy males ($n=5$) and females ($n=6$) with mean age 22.7 years. Fifteen mCi of ^{11}C raclopride were administered over a period of 1 min via intravenous injection while the subject lay in the supine position in the ECAT HR^{+} PET scanner, with head restrained by a custom-molded head holder. PET data were acquired over a 90 min period, with frame durations of 15 s for the first 5 min and 60 s thereafter. Image data were reconstructed with a conventional filtered back projection algorithm, including corrections for attenuation, scatter, dead time, detector nonuniformity and random coincidences. Following reconstruction, data were realigned, frame by frame, to reduce the effect of residual head movement (Alpert et al., 2003). A sum image was formed from the first 17 frames and used in SPM99 to determine the transformation needed to spatially to normalize the data in the MNI coordinate system. These same parameters were used to transform the dynamic PET data, frame-by-frame, to the MNI stereotactic coordinate system.

Parametric images of raclopride binding parameters were computed using the simplified reference region model (SRRM) of Gunn et al. (1997). The SRRM has three parameters: R, k_2 and k_3. Binding potential, BP, was determined as $BP = k_2 / k_3 + R - 1$. Spatially normalized dynamic data sets were read into MATLAB 7 (The Mathworks, Natick, MA). The four dimensional data set was processed as an $m \times N$ array,

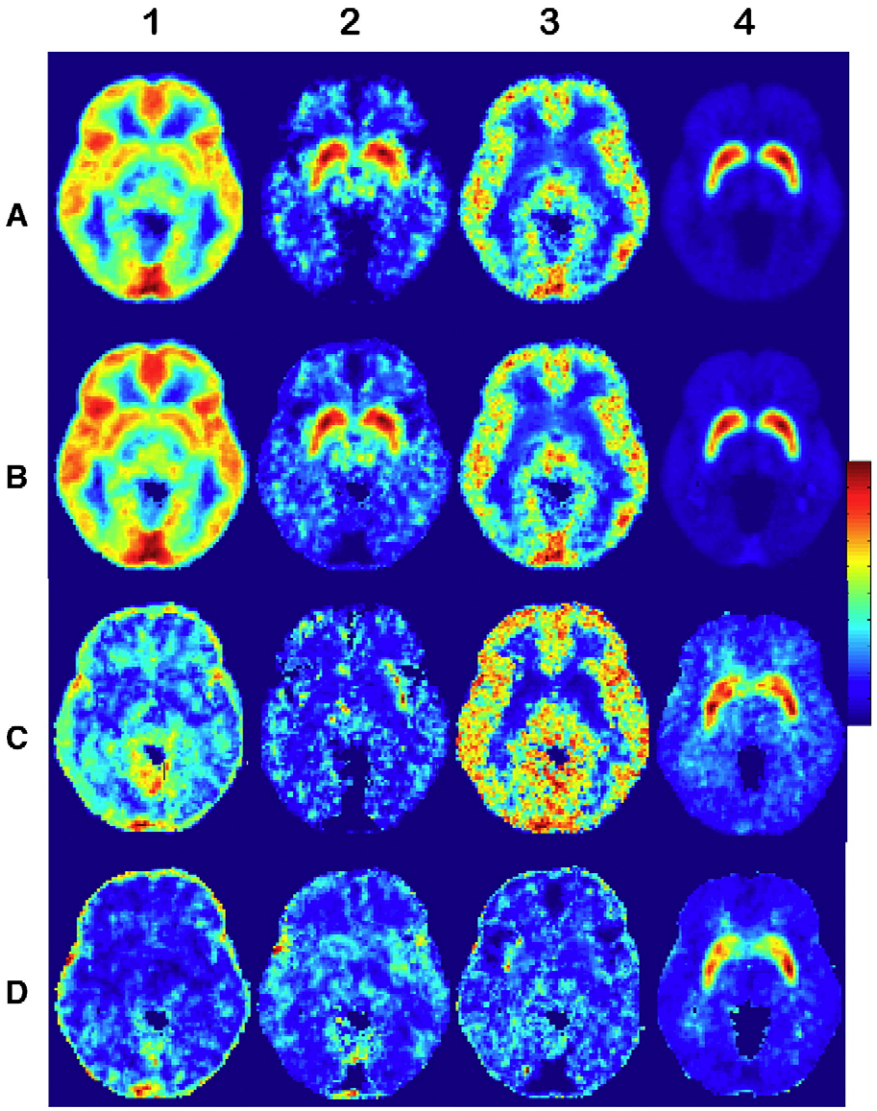

Fig. 1. The table is organized as a four-by-four matrix, with cells labeled by row (A, B, C, D) and column (1, 2, 3, 4) descriptors. The columns depict color maps related to R, k2, k3 and BP. Row 1 and 2 present the parametric images derived by the standard method and Bayesian estimation, respectively; whereas, Rows 3 and 4 present the corresponding maps of standard deviation for each parameter and method. Note the reduction in standard deviation for Bayesian estimation. The relative relationship of the colors is depicted in the color table on the far right of the figure.

where $m=85$ is the number of time points and N is the number of voxels ($128\times128\times63$). The ensemble of 10 parametric images was used to compute the sample mean and standard deviation parametric images. Fig. 1, rows A and C, shows one slice from the mean and standard deviation parametric images of the SRRM.

For this demonstration example, we set the off-diagonal elements of the parameter covariance matrix to zero, $(\Omega_{\boldsymbol{K}}^{-1})_{ij}=\begin{cases}1/\rho_j^2 & i=j\\ 0 & i\neq j\end{cases}$. By choosing S to be a small positive number, one emphasizes the conventional least squares solution and there is little effect from prior measurements but as $S\to\infty$, $\boldsymbol{K}\to\hat{\boldsymbol{K}}$. In our work we varied the value of S in discrete steps over the range 0–20. The criterion for choosing S was to take the smallest value of S that substantially reduced the parameter variances of the cohort. Bayesian regression was computed by a simple modification of an existing nonlinear least squares algorithm. The basic idea is to add r new values to the concentration history, the mean parameter values multiplied by $\sqrt{S\Phi(\boldsymbol{K},0)/\rho_j\hat{k}_j}$, and augment the kinetic model to include the r values of $\sqrt{S\Phi(\boldsymbol{K},0)/\rho_j k_j}$. The results shown below employed the "lsqcurvefit" module in the MATLAB Optimization Toolbox (Version 4.1). No constraints were placed on the parameters during the optimization. All computations were performed voxel-by-voxel on an AMD Opteron personal computer with 8 GB of memory under 64-bit Linux in the MATLAB 7 computational environment. Computation times were long, about 8 h for roughly 30,000 voxel histories.

Bayesian regression was computed for all eleven subjects, based on the priors discussed above. Fig. 2 compares parametric images of binding potential for subject 11 with both standard and Bayesian estimation.

Questions about the degree of bias incurred by Bayesian estimation are not easy to answer. One approach is to compare mean and standard deviation for the cohort using standard and Bayesian estimation. Table 1 presents data for a range of anatomic locations.

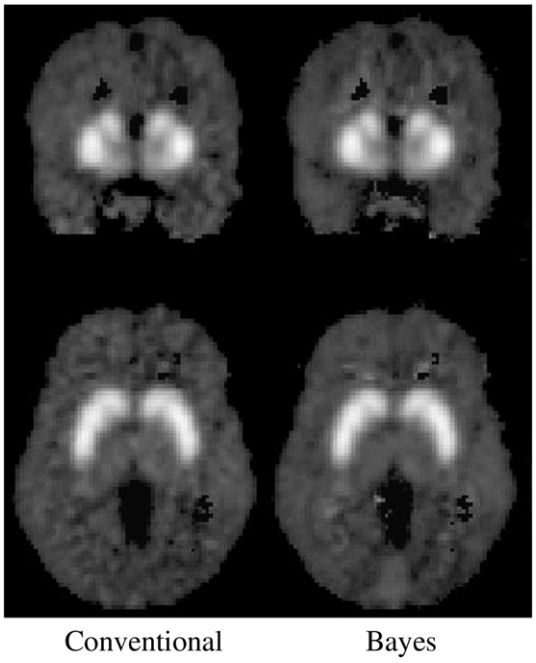

Fig. 2. Illustration of Bayesian estimation. Coronal (top row) and axial (bottom row) projections. Column 1 shows the result of standard estimation of binding potential in subject 11. Column 2 shows the result with Bayesian regression. The images are rendered following a square root transformation to facilitate visualization of both striatal and extrastriatal binding. The Bayesian estimate has increased signal-to-noise ratio with minimal loss of spatial resolution.

Table 1
Binding potential and standard deviation for selected voxels

Anatomic label	X (mm)	Y (mm)	Z (mm)	BP (Gunn)		BP (Bayes)	
				Mean	Std. Dev.	Mean	Std. Dev.
Putamen	−26	4	−6.6	3.68	0.505	3.75	0.465
Caudate	−11	13	−6.6	3.033	0.376	3.075	0.36
Thalamus	−11	−26	0	0.43	0.216	0.433	0.145
Subtantia nigra	4	−29	−33	0.164	0.118	0.169	0.087
Amygdala	35	−18	−11	1.194	0.372	1.288	0.299
Anterior cingulate	11	29	42	0.229	0.178	0.221	0.11
Superior parietal	37	−57	51	0.167	0.123	0.179	0.075
Superior frontal	26	46	24	0.186	0.112	0.225	0.084
MTG	53	2	−24	0.276	0.107	0.286	0.097
IFG	−40	33	−18	0.321	0.147	0.284	0.081
Occipital	9	−92	−13	0.267	0.148	0.312	0.04

Voxel locations X,Y,Z in millimeters from the intersection of the midsagittal plane and the anterior commissure. Anatomic label is nominal. BP refers to voxel-level binding potential estimated by the method of Gunn et al., 1997 or Bayesian regression, as discussed herein.

Fig. 1 (rows B and D) also presents the mean and standard deviation maps for the Bayesian estimates of the parameters and binding potential. The basic result is that means for the parameters are similar for standard and Bayesian estimation but standard deviation is reduced for the Bayesian estimates, particularly in cortical and subcortical regions with low binding potential. For example examination of Table 1 shows that mean binding potential is virtually the same in the thalamus, as estimated by either method, but the standard deviation is about 30% less for Bayesian estimation. Bias is a particular area of concern in situations where the subject under analysis has brain regions that systematically and truly differ from the priors. An example, in the case of raclopride, would be regions with abnormal binding, as might be found in a pathological state. Accordingly, we investigated this case by simulation. We replaced voxel histories in putamen, thalamus, and cortical regions with histories computed to have altered kinetics and binding. Specifically, we generated kinetic curves with the SRRM and to simulate typical noise levels, we added the residuals from real data, obtained from fitting the original voxel curves to the SRRM-generated curves. Two cases were studied, binding potential increased or decreased. Binding potential was increased or decreased by scaling the k_2 parameter of the SRRM by 25%. In the case of the putamen, the binding potential difference was large compared to the variance and the Bayes' term was more significant than for other regions. The results for decreased k_2 parameter are illustrated graphically in Fig. 3 for $S=0$, 4 and 8 and additional details are in Table 2. In simulations of extrastriatal binding the results were similar. In all cases simulated, k_2 and k_3 increased and R_1 decreased, but there was little change in binding potential.

Discussion and conclusions

We have reviewed the theory of Bayesian regression, adding modifications specifically designed for parametric imaging with PET or SPECT. We also conducted a pilot study to evaluate its potential to increase SNR of parametric images of raclopride binding. Our method has its roots in Bayes' theorem, a fundamental pillar of statistical theory. It is general in that it can be applied to any kinetic model and parametric imaging task, provided only that prior information is available about the expected values and statistical distribution of the parameters. In the case of parametric imaging it is advantageous to include the spatial variation in prior parameter values, as this additional information may substantially reduce the random variation in the parameter estimates. It is important to understand why Bayesian regression can be useful in parametric imaging but this subject is complicated by issues unique to dynamic imaging with radionuclides.

The first issue to be discussed is the relation between conventional least squares estimation and Bayesian regression. In conventional

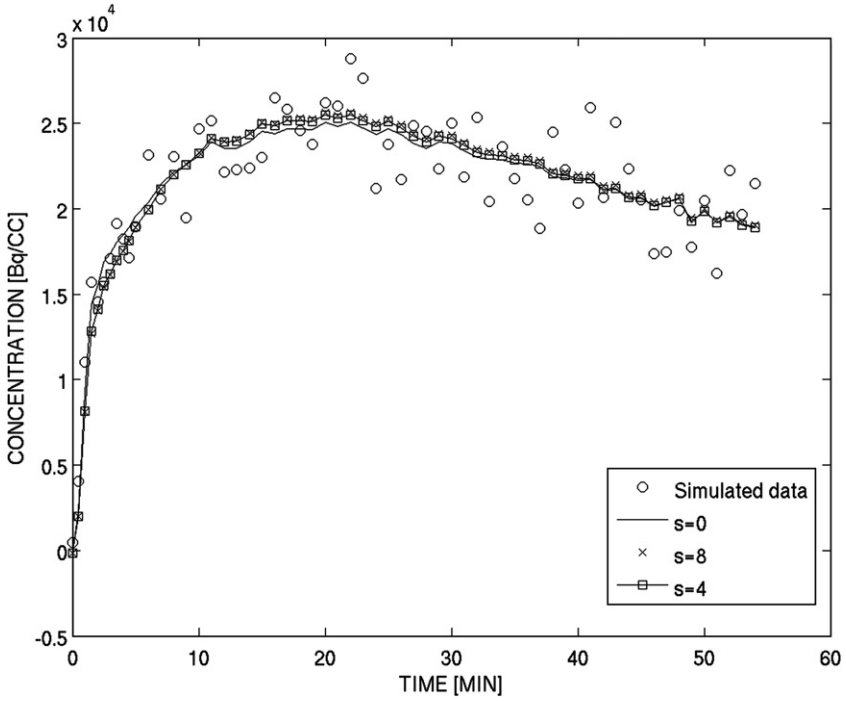

Fig. 3. Simulation study of Bayesian estimation when the data and the prior do not match. The circles are data simulated from a voxel in the putamen with mean (n=10) BP=3.4 and standard deviation 0.2. With data from subject 11, the voxel time-history was analysed with the simplified reference region model and least squares fitting. After decreasing the k_2 value by 25%, the model used to simulate a new time history and the residuals were added to simulate the noise. The solid line represents the new fit with s=0, equivalent to the simplified reference region model prediction. The square and x symbols represent the result of Bayes regression, with s=4 and 8 respectively.

nonlinear least squares estimation, the weighted sum of the squared differences, Ψ, between the model prediction and the measured data is treated as a mathematical function of the parameters. Theoretically, the minimum of Ψ determines a single set of parameters referred to as the least squares estimate. However, because Ψ also depends on the experimental data we are not dealing with a mathematical function, but rather a random variable. Because the measured data may include a substantial amount of random noise, the minimum of the function Ψ can become quite shallow and ill-defined; there may even be more than one local minimum. This is equivalent to saying that, due to noise, there are many sets of parameters that yield model curves that fall within the "cloud" of data points. When the minimum is shallow, many sets of parameters yield similar values of Ψ and consequently the parameters have large variances. Based on prior history, Bayesian regression considers parameter sets for which $\beta = \left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)^T \boldsymbol{\Omega}_{\boldsymbol{K}}^{-1} \left(\boldsymbol{K}-\hat{\boldsymbol{K}}\right)$ is large to be of low probability (see Eq. (4)). In Bayesian regression, another parameter set, with somewhat larger Ψ will be selected if it causes β to be small enough to lower $\Phi(\boldsymbol{K},S)$. In effect, the Bayes' term provides better conditioning for the problem.

A closely related question is: How can Bayesian regression be used to improve parametric imaging? The first problem in applying Bayesian regression is the computation of Eq. (6). The relative magnitude of Ψ and β are usually not well defined in parametric imaging. One expects the magnitude of β to be about one and, theoretically, Ψ should have similar magnitude when the model fits the data. But as we mentioned earlier, Ψ depends on the concentration variance $\boldsymbol{\Omega_C}$ which is a function of position within the image volume. In conventional nonlinear least squares $\boldsymbol{\Omega_C}$ is usually approximated from the number of counts in the projection data, ignoring its spatial variation and the effects of attenuation, scatter and random coincidences. In nonlinear least squares, the absolute magnitude of the concentration variances is not important because it does not affect the minimum of the cost function. Numerous approximations of $\boldsymbol{\Omega_C}$ have been used with success in conventional least squares because the implementation of that algorithm depends only on the "shape" of $\boldsymbol{\Omega_C}$, not its absolute magnitude (c.f. Mazoyer et al., 1986). It is possible to reconstruct quantitative images of $\boldsymbol{\Omega_C}$ when performing filtered back projection reconstruction but that is rarely done (Alpert et al., 1982; Alpert et al., 1991). Currently, there is no standard method for computing $\boldsymbol{\Omega_C}$ when performing iterative reconstructions. To facilitate computation, some standard algorithms for parametric imaging set $\boldsymbol{\Omega_C}$ equal to the identity matrix. Ψ also depends on the scale of the data; that is if more activity is injected Ψ will be proportionately larger. To mitigate this problem we modified Eq. (6), scaling the Bayesian term in the cost function by the product $S\Phi_{\boldsymbol{K}}(\boldsymbol{K},S=0)$, where S is a global strength parameter to be chosen by the investigator. Using this approximation, S becomes independent of position in the image volume while $\Phi_{\boldsymbol{K}}(\boldsymbol{K},S=0)$ is computed at each voxel, retaining the desirable features of Bayesian regression. The tradeoff is that we cannot claim that the estimates are optimal in a statistical sense and S must be chosen based on empirical criteria, such as we described in the demonstration example or perhaps by Monte Carlo simulation.

The next complication is that we expect that parametric images will exhibit spatial patterns associated with the underlying anatomy and physiology. A method is needed to match the spatial patterns in the prior information to the data at hand. A complete solution to this problem is not available, but tractable approximations can be

Table 2
Simulation study: BP differs from prior by 25%

S	Parameters			
	R_1	k_2	k_3	BP
Simulation	1.279	0.165	0.0700	2.638
0	1.262	0.170	0.0719	2.630
4	1.092	0.202	0.0800	2.612
8	1.080	0.204	0.0795	2.639

Note: The priors for this simulation were R_1: 1.079±0.282, k_2: 0.218±0.025 and k_3: 0.637±0.011.

made in brain imaging. In our example of raclopride binding studies, we used a cohort of normal subjects studied in previous work (Badgaiyan et al., 2007). Within the limitations of spatial normalization, it is possible to transform an individual's parametric images to a standard orientation, size and position in space. Spatial normalization is not perfect. It is well known that there is residual anatomic and functional variability, particularly in some cortical areas (Rajkowska and Goldman-Rakic, 1995; Thompson et al., 1996, 1997; Evans et al. 1997; Grachev et al. 1999). Once the transformations are effected, it is a simple matter to form sample mean and standard deviation or covariance images for the cohort's parametric images.

There is another consideration and it may be the most important of all. We must decide what prior information to use. What should the characteristics of the prior distribution be when one is studying subjects with known or expected pathology? It is clear that Bayesian estimation will have the largest influence when analyzing voxel histories that are very noisy, the ones for which conventional least squares estimates are, in a sense, ambiguous. But it does this by biasing the result toward the expected value. This is likely to be acceptable in normal cohorts or fairly homogeneous control groups. It may be more controversial in situations where a decision about a single individual is wanted (Bromiley et al., 2003). Priors incorporating the probability of regional abnormalities may be difficult to construct because they may require very large sample size. This is an area that may be illuminated by future research with Monte Carlo simulation. Even when the choice of prior is clear cut, it is important that the cohort of priors and data undergoing Bayesian regression be very similar. The investigators should also ensure that the new subjects match the priors in terms of activity injected, acquisition protocol, reconstruction details and scanner characteristics. Otherwise, the underlying assumptions of Bayesian regression will be violated, the distribution of the parameter values will differ from those measured in the new experiment and the estimates may be biased in ways that cannot be anticipated.

In our pilot study we used ten subjects to form the prior. Ideally a much larger cohort should be used. In this pilot study, we had to assume that the voxel data followed Gaussian statistics as described by their mean and variance. If larger cohorts of prior data are available, one can test the hypothesis about Gaussian distributions or fall back on the empirical distributions mined from the cohort. It should be noted that prior parameter distributions formed by stereotactic averaging include both statistical and anatomic variability. The anatomic variability may be a significant problem in some brain regions, possibly making the results invalid. In the pilot study, distortions were noted in some areas with large admixtures of white and gray matter. On the other hand, once these regions are identified, they can be eliminated from consideration and conventional estimation techniques can be used for those voxels.

The methodology employed in the pilot study was not ideal. The demonstration would have been more rigorous and credible had we had analyzed a large prior population and a correspondingly large test population. As it is, we used the sample population to form the priors and evaluate bias and noise reduction. However, the analysis of a single subject not included in the prior population also showed the expected noise reduction without obvious increase in bias. When analyzed in this way, the same population analyzed by nonlinear least squares and our Bayesian regression exhibited significant noise reduction. Standard deviation of grey matter binding potentials decreased by at least 25%.

We noted earlier in this paper that some investigators had used penalized least squares to reduce the noise propagation in parametric imaging. Bayesian regression is also a form of penalized least squares and, so, it is fair to ask how it differs from techniques such as ridge regression? An examination of the operational equations for ridge regression and Bayesian regression shows that they are formally identical! In the method described by Zhou et al., the penalty term $(\boldsymbol{K}-\hat{\boldsymbol{K}})^T W^{-1}(\boldsymbol{K}-\hat{\boldsymbol{K}})$ is approximated by estimating $\hat{\boldsymbol{K}}$ and $\boldsymbol{\Omega}$ for a given voxel by averaging over nearby, surrounding voxels. Estimating $\boldsymbol{\Omega}$ by sample statistics on nearby voxels assumes that the voxel data are statistically independent, but this is not true for reconstructed data. An advantage of the ridge regression method is that it requires no additional data. For better, or worse, $\hat{\boldsymbol{K}}$ and $\boldsymbol{\Omega}^{-1}$ are determined from the data at hand. The improvement in SNR depends entirely on the local averaging employed in the calculation of $\hat{\boldsymbol{K}}$ and $\boldsymbol{\Omega}^{-1}$. The disadvantage for ridge regression is that averaging over a larger volume gains SNR at the expense of resolution and bias.

Our data do not provide a definitive conclusion about the bias introduced by Bayesian regression. Therefore, bias should be assessed whenever new applications are considered. Table 1 lists binding potential, computed by Bayesian and standard regression, for a sampling of cortical and subcortical regions. The percent differences (standard-Bayes) in the mean of the model parameters is 4.4% for BP. Data on the differences by method for the other model parameters (not shown for) the other model parameters, R, k_2 and k_3 are similar: 2.5% for R, −4.7% for k_2 and 7.2% for k_3. Because the parametric images are noisy, these percent differences should be interpreted with care since the standard errors of the mean are of the same order of magnitude as the differences.

In the pilot study of raclopride binding parameters comparing standard and Bayesian estimation, the SNR was, as expected, increased with Bayesian estimation. In our implementation we used a “Bayes parameter” $S=8$, based on empirical observation. We found that the improvements in SNR depended on the local noise level in the standard least squares estimation procedure. Areas of lower SNR benefited most; whereas, there were more modest improvements, for example, in the striatum. This finding can be understood by considering the sources of noise in the formation of the image cohorts: These include the statistical fluctuations in the voxel-level concentrations values, the functional variations from subject to subject, and the residual anatomic variation after stereotactic normalization. This variability is least in the striatum and highest in cortical grey matter and goes some way in explaining our findings.

Our method of parametric imaging with Bayesian regression has a unique advantage. It retains the original resolution of the study. It does not employ smoothing in any direct way, neither spatially nor temporally. It improves the SNR of the estimated parameters by favoring solutions that are in accord with the prior distribution of parameters at each location in stereotactic space. There is, however, no free lunch. Improved SNR involves choosing $S>0$. In the limit of very large values of S, the parametric images become identical to the sample mean of the prior distribution. A reasonable choice of S involves empirical study or realistic simulation. In the example of raclopride binding potential, we see that there is about 10 to 1 difference between cortical grey matter and striatum. As one would expect, the biggest improvement in SNR is in the noisiest regions. In its present state of development, and considering the caveats discussed above, the method is best suited for research studies which employ groups of normal subjects. Under such conditions, our method of Bayesian regression is capable of significant increase in the signal-to-noise ratio of parametric images.

Acknowledgment

Dr. Yuan was supported by the Postgraduate Program in Radiological Science, 5T32EB002102, a National Institute of Biomedical Imaging and Bioengineering postdoctoral training grant.

References

Alpert, N.M., Chesler, D.A., Correia, J.A., Ackerman, R.H., Chang, J.Y., Finklestein, S., Davis, S.M., Brownell, G.L., Taveras, J.M., 1982. Estimation of local statistical noise in emission computed tomography. IEEE Trans Med Imag MI-1 142–146.

Alpert, N.M., Barker, W.C., Gelman, A., Weise, S., Senda, M., Correia, J.A., 1991. The precision of positron emission tomography: theory and measurement. J. CBF Metab. 11, A26–A30.
Alpert, N.M., Badgaiyan, R.D., Livni, E., Fischman, A.J., 2003. A novel method for noninvasive detection of neuromodulatory changes in specific neurotransmitter systems. NeuroImage 19, 1049–1060.
Alpert, N.M., Reilhac, A., Chio, T.C., Selesnick, I., 2006. Optimization of dynamic measurement of receptor kinetics by wavelet denoising. NeuroImage 30, 444–451.
Badgaiyan, R.D., Fischman, A.J., Alpert, N.M., 2007. Striatal dopamine release in sequential learning. NeuroImage 38, 549–556.
Bard, Y., 1974. Nonlinear Parameter Estimation. Academic Press, New York.
Bevington, P.R., Robinson, D.K., 2003. Data Reduction and Error Analysis for the Physical Sciences. McGraw-Hill.
Bromiley, P.A., Thacker, N.A., Scott, M.L.J., Pokric, M., Lacey, A.J., Cootes, T.F., 2003. Bayesian and non-Bayesian probabilistic models for medical image analysis. Image Vis. Comput. 21, 851–864.
Chen, B., Huang, S.-C., Hawkins, R.A., Phelps, M.E., 1988. An evaluation of Bayesian regression for estimating oxygen utilization with oxygen-15 and dynamic PET. IEEE Trans. Med. Imag. 7, 257–263.
Collins, D.L., Peters, T.M., Evans, A.C., 1994. An automated 3D non-linear image deformation procedure for determination of gross morphometric variability in the human brain. Proc. Vis. Biomed. Comp. (SPIE) 3, 180–190.
Cselenyi, Z., Olsson, H., Farde, L., Gulyas, B., 2002. Wavelet-aided parametric mapping of cerebral dopamine D2 receptors using the high affinity PET radioligand [^{11}C]FLB 457. NeuroImage 17, 47–60.
Evans, A.C., Collins, D.L., Holmes, C.J., et al., 1997. A 3D probabilistic atlas of normal human neuroanatomy. Proc. 3rd Int. Conf. Funct. Mapp. Hum. Brain, S349.
Grachev, I.D., Berdichevsky, D., Rauch, S.L., Heckers, S., Kennedy, D.N., Caviness, V.S., Alpert, N.M., 1999. A method for assessing the accuracy of intersubject registration of the human brain using anatomic landmarks. NeuroImage 9, 250–268.
Gunn, R.N., Lammertsma, A.A., Hume, S.P., Cunningham, V.J., 1997. Parametric imaging of ligand–receptor binding in PET using a simplified reference region model. NeuroImage 6, 279–287.
Kimura, Y., Hsu, H., Toyama, H., Senda, M., Alpert, N.M., 1999. Improved signal-to-noise ratio in parametric images by cluster analysis. NeuroImage 9, 554–561.
Kimura, Y., Naganawa, M., Yamaguchi, J., Takabayashi, Y., Uchiyama, A., Oda, K., Ishii, K., Ishiwata, K., 1996. MAP-based kinetic analysis for voxel-by-voxel compartment model estimation: detailed imaging of the cerebral glucose metabolism using FDG. NeuroImage 29, 1203–1211.
Mazoyer, B.M., Huesman, R.H., Budinger, T.F., Knittel, B.L., 1986. Dynamic PET data analysis. J. Comput. Assist. Tomogr. 10, 645–653.
Millet, P., Ibanez, V., Delforge, J., Pappata, S., Guimon, J., 2000. Wavelet analysis of dynamic PET data: application to the parametric imaging of benzodiazepine receptor concentration. NeuroImage 11, 458–472.
Muzic Jr, R.F., Christian, B.T., 2006. Evaluation of objective functions for estimation of kinetic parameters. Med. Phys. 33, 342–353.
O'Sullivan, F., Saha, A., 1999. Use of ridge regression for improved estimation of kinetic constants from PET data. IEEE Trans. Med. Imag. 18, 115–125.
Rajkowska, G., Goldman-Rakic, P., 1995. Cytoarchitectonic definition of pre-frontal areas in the normal human cortex: II. Variability in locations of areas 9 and 46 and relationship to the talairach coordinate system. Cereb. Cortex 5, 323–337.
Shidahara, M., Ikoma, Y., Kershaw, J., Kimura, Y., Naganawa, M., Watabe, H., 2007. PET kinetic analysis: wavelet denoising of dynamic PET data with application to parametric imaging. Ann. Nucl. Med. 21, 379–386.
Thompson, P.M., Schwartz, C., Lin, R.T., Khan, A.A., Toga, A.W., 1996. 3D statistical analysis of sulcal variability in the human brain. J. Neurosci. 16, 4261–4274.
Thompson, P., MacDonald, D., Mega, M.S., et al., 1997. Quantifying and correcting for variable cortical morphology in functional imaging using a deformable probabilistic brain atlas. Hum. Brain Mapp. 4, 5423.
Turkheimer, F.E., Aston, J.A., Banati, R.B., Riddell, C., Cunningham, V.J., 2003. A linear wavelet filter for parametric imaging with dynamic PET. IEEE Trans. Med. Imag. 22, 289–301.
Yu, X., White, L.T., Doumen, C., Damico, L.A., LaNoue, K.F., Alpert, N.M., Lewandowski, E. D., 1995. Kinetic analysis of dynamic ^{13}C NMR spectra: metabolic flux, regulation, and compartmentation in hearts. Biophys. J. 69, 2090–2102.
Zhou, Y., Huang, S.-C., Bergsneider, M., 2001. Linear ridge regression with spatial constraint for generation of parametric images in dynamic positron emission tomography studies. IEEE Trans. Nucl. Sci. 48, 125–130.
Zhou, Y., Huang, S.C., Bergsneider, M., Wong, D.F., 2002. Improved parametric image generation using spatial–temporal analysis of dynamic PET studies. NeuroImage 15, 697–707.
Zhou, Y., Endres, C.J., Brasic, J.R., Huang, S.C., Wong, D.F., 2003. Linear regression with spatial constraint to generate parametric images of ligand–receptor dynamic PET studies with a simplified reference tissue model. NeuroImage 18, 975–989.