

## MIXTURE MODELING FOR DYNAMIC PET DATA

Huiping Jiang and R. Todd Ogden

*New York State Psychiatric Institute and Columbia University*

*Abstract:* Standard kinetic modeling of dynamic positron emission tomography (PET) data requires specifying a compartment structure and fitting the appropriate kinetic model using nonlinear least squares algorithms separately for each voxel in the brain. This approach is not completely satisfactory because of a natural reluctance researchers have to specifying a particular compartmental model to be applied to all voxels and, in addition, there are parameter identifiability issues for all but the simplest models. This paper presents new methodology for modeling dynamic PET data that works by “borrowing strength” across all voxels, expressing each voxel’s data as a linear combination of a small number of components that are estimated from the data. Though based on a kinetic modeling structure, it does not require a choice of compartmental system and allows for data-adaptive choice of model order. The spatial autocorrelation throughout the brain is modeled with a conditional autoregressive (CAR) model. Estimation of model parameters is accomplished through iterative optimization based on nonlinear weighted least squares, and selection of the number of components is based on a modified information criterion. This methodology may be applied either at a voxel-level or in a region of interest (ROI) analysis. Performance of the method is evaluated with simulated and real data.

*Key words and phrases:* Conditional autoregressive modeling, kinetic modeling, voxel.

### 1. Introduction

One use of positron emission tomography (PET) technology is to determine the distribution of a target neuroreceptor throughout the brain. Such studies can help in the investigation of the neuropathology of psychiatric and neurologic illnesses.

In a typical study, a radioactive ligand developed specifically for the receptor under investigation is injected into a subject’s bloodstream, and the concentration of the ligand is measured over time using the PET imaging modality. Each ligand has its own distinct kinetic behavior *in vivo*, and modeling these kinetics is the key to estimation of the density of the target receptor in each location throughout the brain.

A popular general approach to describe such data is compartmental modeling. Typically, a model is fit separately to the time-activity curve (TAC) for each

voxel (or region of interest (ROI)). This results in an image of estimated kinetic parameters which, once properly registered, may then be compared across subjects. This requires that researchers specify a particular compartmental structure and, as may be expected, results are heavily dependent on the particular model selected. Also, for many ligands, parameters from all but the simplest of kinetic models are not identifiable when fit to each voxel separately. Our approach is to estimate kinetic parameters simultaneously across voxels (or across ROIs) in order to allow for more complex kinetic structure.

## 2. Kinetic Model with Plasma Input

A common approach to analyze data from PET imaging studies is to fit compartmental models using a plasma “input function” (Mintun et al. (1984)). The input function specifies the concentration of the free tracer in the plasma over time, and models using this input function require that samples from an arterial line be measured over time during the imaging study. Upon entering the brain, the ligand particles “drift” among several states of activity (“compartments”) — e.g., bound to the target receptor, bound to another receptor, or “free”. The kinetic parameters specify the rate at which particles “move” from one state to another. This model is sketched in Figure 1. For each compartment  $k$  ( $1 \leq k \leq K$ ), the concentration of the ligand at time  $t$  is denoted by  $C_k(t)$ . Concentration for the plasma compartment (to be modeled independently of the brain data modeling) is given by  $C_p(t)$ . The PET imaging modality is unable to distinguish among the various compartments in the brain, and therefore only the total brain

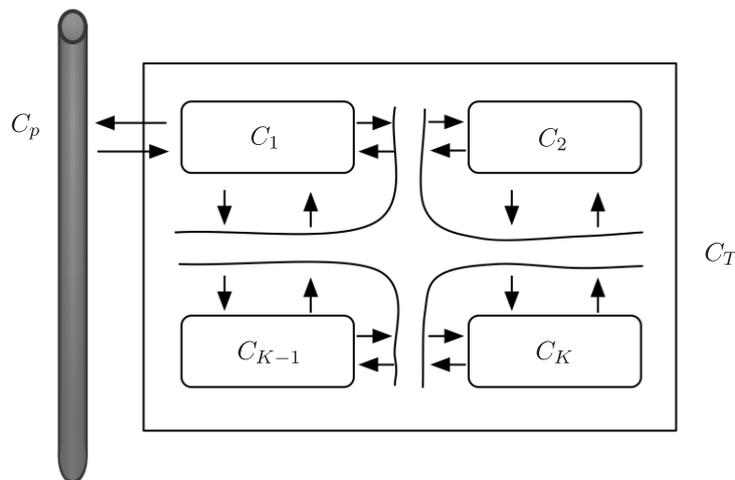


Figure 1. Generalized tissue model as in Gunn et al. (2001)

concentration  $C_T(t) = \sum_{k=1}^K C_k(t)$  can be measured at each voxel. The input function  $C_p(t)$  is measured from the arterial samples drawn during imaging and modeled separately from the PET data.

If transit rates among the  $K$  brain compartments and the plasma compartment are constant over time, then the model may be represented by a linear compartmental system defined in term of its state space representation (Gunn, Gunn and Cunningham (2001)). Defining  $\mathbf{x}(t) = (C_1(t), \dots, C_K(t))'$  and  $\dot{\mathbf{x}}(t) = ((d/dt)C_1(t), \dots, (d/dt)C_K(t))'$ , the system is given by

$$\dot{\mathbf{x}}(t) = A\mathbf{x}(t) + \mathbf{b}C_p(t), \quad \mathbf{x}(0) = \mathbf{0}. \tag{2.1}$$

Here,  $A$  is the  $K \times K$  state transition matrix made up of linear combinations of the rate constants describing the transfer of material between compartments, and  $\mathbf{b} = (r, 0, \dots, 0)'$  is a  $K$ -vector.

At time  $t$ , the total brain concentration is  $C_T(t) = \mathbf{1}'\mathbf{x}(t)$ . The general solution for (2.1) is given by

$$C_T(t) = \sum_{k=1}^K \beta_k \left( e^{-\gamma_k \cdot} \otimes C_p \right) (t),$$

where  $\otimes$  is the convolution operator. The PET macroparameters to be estimated are the  $\gamma_k$ 's and  $\beta_k$ 's with the restriction that  $\beta_k \geq 0$  for all  $k = 1, \dots, K$ . Each element of the transition matrix  $A$  and of the vector  $\mathbf{b}$  is a function of the  $\gamma_k$ 's and  $\beta_k$ 's. For instance, for  $K = 1$ , we have  $A = \gamma_1$  and  $b = \beta_1$ .

A common outcome measure estimated through this type of modeling is the "total volume of distribution" ( $V$ ), which is equal to the integral of the "impulse response function":

$$V = \int_0^\infty \sum_{k=1}^K \beta_k e^{-\gamma_k t} dt = \sum_{k=1}^K \frac{\beta_k}{\gamma_k}. \tag{2.2}$$

In (2.1),  $C_T(t)$  is a continuous measurement but, in practice, the observation  $C_T(t_i)$ ,  $i = 1, \dots, T$ , corresponding to the tomographic signal, is discrete and measures the time-averaged value of  $C_T(t)$  over the time interval  $(t_i^s, t_i^e)$ , i.e.,

$$C_T(t_i) = \frac{1}{t_i^e - t_i^s} \int_{t_i^s}^{t_i^e} C_T(t) dt.$$

where  $t_i = (t_i^e + t_i^s)/2$ ,  $t_1^s = 0$  and  $t_{i+1}^s = t_i^e$ . Observations taken from this system are contaminated with additive noise. It is customary to model PET data contaminated with additive noise as

$$y(t_i) = C_T(t_i) + \epsilon(t_i) = \sum_{k=1}^K \beta_k \left( e^{-\gamma_k \cdot} \otimes C_p \right) (t_i) + \epsilon(t_i), \quad i = 1, \dots, T, \tag{2.3}$$

where the measurement error,  $\epsilon(t_i)$ , is taken to be uncorrelated with  $E(\epsilon(t_i)) = 0$ , and  $\text{var}(\epsilon(t_i), \epsilon(t_j)) = \sigma^2(t_i)$ . As the measurements in PET are actually the average of counts acquired during scanning intervals, the variances of the measurement noise are inversely proportional to the lengths of the time intervals (Chen, Huang and Yu (1991)). The general variance structure of PET measurement can be described as

$$\sigma^2(t_i) = \frac{g(C_T(t_i), t_i)}{w(t_i)},$$

where  $g(\cdot)$  is a function with positive value, and  $w(t_i) = t_i^e - t_i^s$ . Without loss of generality, we assume  $g(C_T(t_i), t_i)$  is a constant value, i.e.,  $g(C_T(t_i), t_i) = \sigma^2$  and  $\sigma^2(t_i) = \sigma^2/w(t_i)$ . We note that many choices for the weight function are in common use in kinetic modeling with PET data (see, e.g., Cobelli, Foster and Toffolo (2000), and references therein). Our development throughout is appropriate for any choice of weight function.

As mentioned earlier,  $C_p$  may be estimated separately from the PET data using measured plasma concentrations. Given this estimate of  $C_p$ , the  $\beta_k$ 's and  $\gamma_k$ 's may be estimated from the PET data by "plugging in" the estimate  $\hat{C}_p$  for  $C_p$  in (2.3). (We henceforth assume that  $C_p$  has been estimated in such a way, and thus regard  $C_p$  as "fixed" and "known". Resulting inference on kinetic parameters is understood to be conditional on the measured plasma data.) Thus in standard kinetic modeling with an input function, parameters are estimated by minimizing

$$\sum_{i=1}^T w(t_i) \left( y(t_i) - \sum_{k=1}^K \beta_k \left( e^{-\gamma_k \cdot} \otimes C_p \right) (t_i) \right)^2$$

over all choices of  $\beta_1, \dots, \beta_K, \gamma_1, \dots, \gamma_K$  using a nonlinear least squares algorithm.

### 3. Basis Function Methods

Kinetic modeling as described in Section 2 is applied frequently in practice but, in many applications, alternative approaches that do not require specifying a particular compartmental structure would be favored. The so-called basis function methods do not require specification of compartmental structure (nor even the assumption that compartmental structure is the same for all voxels) and also have the advantage of computational efficiency (kinetic models are typically fit by nonlinear least squares at relatively high computational cost). This approach involves creating a library of "basis functions" of the form  $(e^{-\gamma_k \cdot} \otimes C_p)(t)$  for a range of  $\gamma_k$  values. The most challenging problem encountered in application of such methods is determining which basis functions are needed for each

voxel (the model selection problem); once this has been determined, estimation of coefficients is a linear regression problem.

The library of basis functions must be large enough to allow for a wide range of kinetic behavior across voxels, but for stable estimation, a relatively sparse subset of basis functions must be chosen. In the spectral analysis method introduced by Cunningham and Jones (1993), sparsity is ensured by constraining the coefficients to be nonnegative. Gunn et al. (2002) address the sparsity issue by adding an  $L_1$  penalty to the coefficient vector in the optimization.

These methods are applied to data by fitting each voxel separately; that is, the data from each voxel determine the choice of basis functions to be included in the fit as well as the coefficients of the “selected” basis functions.

#### 4. A Mixture Formulation

We offer an alternative approach to the basis function methods in which considerably fewer “basis” functions are used to fit data for each voxel, and so that these functions are determined from all voxels simultaneously. We begin by observing that if the “true”  $\gamma_k$  parameters were known, each function

$$f(t_i|\gamma_k) = \left( e^{-\gamma_k \cdot} \otimes C_p \right) (t_i) \quad (4.1)$$

may be regarded as a single component function, and the model may be written as

$$C_T(t_i) = \sum_{k=1}^K \beta_k f(t_i|\gamma_k).$$

Thus, with a pre-specified set of  $K$  such component functions, the coefficients  $\{\beta_k\}$  could be estimated using standard nonnegative linear least squares (NNLS) (Lawson and Hanson (1974)).

In practical application, of course, the  $\gamma_k$  values will not be known and therefore must be estimated from the data. In a departure from the basis function approaches, we propose to fit all voxels at once, requiring that the  $K$  component functions be the same across all voxels.

To accomplish this, we incorporate a mixture representation, following the general approach taken by O’Sullivan (2006), but requiring that component functions take the parametric form (4.1). Since we fit all voxels simultaneously, we also incorporate a spatial model to account for correlation between voxels.

Here we lay out our model for the PET data which is based on the requirement that the  $\gamma_k$  values be the same for all voxels but the coefficients may vary from voxel to voxel. Thus, the model for the  $j$ th voxel may be written as

$$y_j(t_i) = \sum_{k=1}^K \beta_{jk} f(t_i|\gamma_k) + \frac{\epsilon_j(t_i)}{\sqrt{\nu_j}}, \quad j = 1, \dots, N, \quad i = 1, \dots, T, \quad (4.2)$$

with  $\prod_{j=1}^N \nu_j = 1$  and

$$\text{cov}(\epsilon_j(t_i), \epsilon_k(t_l)) = \frac{\sigma^2 \rho_{jk} \delta(t_i - t_l)}{\sqrt{w(t_i)w(t_l)}}.$$

Here,  $\nu_j > 0$ ,  $j = 1, \dots, N$ , is a parameter to adjust the difference in variance among voxels;  $N$  is the total number of voxels in the brain;  $\delta(\cdot)$  is the delta function; and  $\rho_{jk}$  represents the spatial correlation between measurements made on voxel  $j$  and voxel  $k$  (thus  $\rho_{jj} = 1$  for all  $j$ ). When fitting individual voxels, it might be reasonable to assume that the variance is constant, i.e.,  $\nu_j = 1$  for all  $j = 1, \dots, N$ . However, when this methodology is to be applied to ROI data, the observations are summary statistics (typically, simple averages) of the concentrations of voxels in the anatomically defined region. Since ROIs differ substantially in terms of size (number of voxels), it is necessary to allow for different variances across ROIs (e.g.,  $\nu_j \propto n_j$ , where  $n_j$  is the number of voxels in the  $j$ th ROI). The algorithm presented here also allows for estimation of the  $\nu_j$ 's.

Let  $\mathbf{Y}_j$  represent the  $j$ th TAC in the brain for  $j = 1, \dots, N$ :

$$\mathbf{Y}_j = (y_j(t_1), y_j(t_2), \dots, y_j(t_T))'.$$

The  $T \times K$  matrix  $F_\gamma$  contains the  $K$  component functions along its columns:

$$F_\gamma = \begin{pmatrix} f(t_1|\gamma_1) & f(t_1|\gamma_2) & \cdots & f(t_1|\gamma_K) \\ f(t_2|\gamma_1) & f(t_2|\gamma_2) & \cdots & f(t_2|\gamma_K) \\ \vdots & \vdots & \ddots & \vdots \\ f(t_T|\gamma_1) & f(t_T|\gamma_2) & \cdots & f(t_T|\gamma_K) \end{pmatrix}.$$

Let  $\gamma = (\gamma_1, \dots, \gamma_K)'$ , and  $\beta_j$  be the vector of coefficients for the  $j$ th voxel:  $\beta_j = (\beta_{j1}, \beta_{j2}, \dots, \beta_{jK})'$ . Then we can express (4.2) in matrix form as

$$\mathbf{Y}_j = F_\gamma \beta_j + \nu_j^{-\frac{1}{2}} \mathbf{E}_j = F_\gamma \beta_j + \nu_j^{-\frac{1}{2}} W^{-\frac{1}{2}} \mathbf{U}_j,$$

with  $\mathbf{E}_j = (\epsilon_j(t_1), \dots, \epsilon_j(t_T))'$ ,  $\mathbf{U}_j = (u_j(t_1), \dots, u_j(t_T))'$ ,  $W = \text{diag}(w(t_1), \dots, w(t_T))$ ,  $\mathbf{E}(\mathbf{U}_j) = \mathbf{0}_{T \times 1}$ , and  $\text{Var}(\mathbf{U}_j) = \sigma^2 I_T$ . This can also be expressed as

$$\sqrt{\nu_j} W^{\frac{1}{2}} \mathbf{Y}_j = \sqrt{\nu_j} W^{\frac{1}{2}} F_\gamma \beta_j + \mathbf{U}_j.$$

Defining  $\mathbf{Z}_j = \sqrt{\nu_j} W^{1/2} \mathbf{Y}_j$ ,  $R_\gamma = W^{1/2} F_\gamma$ ,  $\beta_j^* = \sqrt{\nu_j} \beta_j$ , and  $B^* = [\beta_1^* | \beta_2^* | \dots | \beta_N^*]$ , the model may be expressed concisely as

$$\mathbf{Z} = R_\gamma B^* + \mathbf{U},$$

where  $Z = [Z_1|Z_2|\dots|Z_N]$  and  $U = [U_1|U_2|\dots|U_N]$ .

If we assume that the  $T$  rows of  $U$  are iid multivariate normal random vectors with mean  $\mathbf{0}_N$  and variance covariance matrix  $\Sigma$  with  $(i, j)$  element equal to  $\rho_{ij}\sigma^2$ , then the likelihood function for the data can be expressed (see, e.g., Rencher (2002)):

$$L(B^*, \gamma, \Sigma|Z) = (2\pi)^{-\frac{TN}{2}} |\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr} \left( (Z - R_\gamma B^*) \Sigma^{-1} (Z - R_\gamma B^*)' \right) \right\}.$$

For a given value of  $R_\gamma$  and  $\Sigma$ , the likelihood is maximized at  $\hat{B}^* = \arg \min_{\{\beta_k\}} \text{tr}((Z - R_\gamma B^*) \Sigma^{-1} (Z - R_\gamma B^*)')$ . Since  $\{\beta_k\}$  has to be nonnegative  $\hat{B}^*$  can be obtained by fitting separately for each TAC in the brain using the NNLS method. It remains to maximize the profile likelihood

$$L(\gamma, \Sigma, \hat{B}^*|Z) = (2\pi)^{-\frac{TN}{2}} |\Sigma|^{-\frac{T}{2}} \exp \left\{ -\frac{1}{2} \text{tr}(\Sigma^{-1} \hat{U}' \hat{U}) \right\} \tag{4.3}$$

over  $\gamma$  and  $\Sigma$ , where  $\hat{U}$  is the matrix of residuals:  $\hat{U} = Z - R_\gamma \hat{B}^*$ .

We model the covariance structure  $\Sigma$  among voxels using a conditional autoregressive (CAR) model. This model is useful in this situation because it involves only one parameter to be estimated, and because the computationally simple form for the inverse of the covariance matrix makes maximum likelihood estimation feasible even for very large data sets. This CAR model is similar in spirit to the autoregressive model considered by O’Sullivan (2006) and Maitra and O’Sullivan (1998). In the CAR model, the conditional expectation of  $Z_j$  given the observations at its neighbors is  $R_\gamma \beta_j^*$  plus a weighted sum of the mean-centered responses of its neighbors. Defining  $Z_{-j}$  to be the matrix resulting from eliminating the  $j$ th column of  $Z$ , we model the data as, for  $j = 1, \dots, N$ ,

$$E(Z_j|Z_{-j}) = R_\gamma \beta_j^* + \rho \sum_{i=1}^N \phi_{ji} (Z_i - R_\gamma \beta_i^*), \quad j = 1, \dots, N;$$

$$\text{Var}(Z_j|Z_{-j}) = \sigma^2 I_T,$$

The parameter  $\rho$  determines the magnitude of the spatial neighborhood effect, and the  $\phi_{ji}$  are weights that determine the relative influence of location  $i$  on the location  $j$  with  $\phi_{jj} = 0$  for all  $j$ . Thus the  $i$ th row of  $Z$  has variance

$$\sigma^2 (I_N - \rho \Phi)^{-1}, \tag{4.4}$$

where  $\Phi = \{\phi_{ij}\}_{N \times N}$  is the *proximity matrix*. The spatial correlation structure is expressed in terms of  $\Phi$ . One common choice is to define a neighborhood

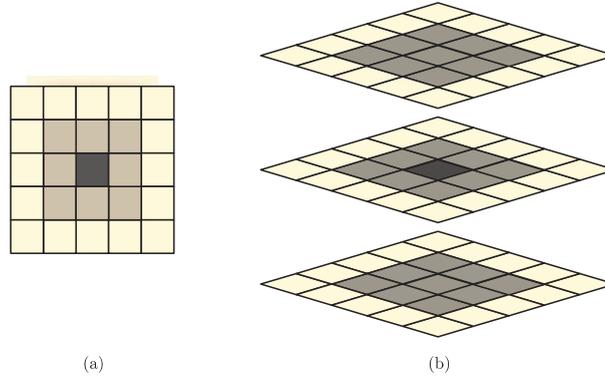


Figure 2. Illustration of spatial connectedness (contiguity) for a regular lattice for (a) 2D space and (b) 3D space.

about each voxel and set  $\phi_{ij} = \phi_{ji} = 1$  if  $i$  and  $j$  are in the same neighborhood (and 0 otherwise). In a typical 2-D lattice, one voxel has 8 neighbors, while a voxel has 26 neighbors in a typical 3-D lattice (Figure 2). In practice, one can select an appropriate weight function based on the model fit and how well the model accounts for autocorrelation in the residuals. Requiring the variance-covariance matrix in (4.4) to be non-singular imposes constraints on  $\Phi$  and  $\rho$ . If  $\theta_{max}$  and  $\theta_{min}$  are the largest and smallest eigenvalues of  $\Phi$ , and if  $\theta_{min} < 0$  and  $\theta_{max} > 0$ , then  $1/\theta_{min} < \rho < 1/\theta_{max}$  (Haining (1990)). Often, the elements of  $\Phi$  are standardized by dividing each entry in  $\Phi$  by its row sum, in which case  $\theta_{max}=1$ . In brain imaging applications, it is reasonable to assume the spatial neighborhood effect,  $\rho$ , satisfies  $0 \leq \rho < 1$ . With this parameterization of  $\Sigma$ , the profile likelihood function (4.3) becomes  $L(\gamma, \sigma^2, \rho|Z)$ . This is maximized with respect to  $\sigma^2$  at  $\hat{\sigma}^2 = (TN)^{-1}\text{tr}((I - \rho\Phi)\hat{U}'\hat{U})$ , yielding the profile log-likelihood

$$\log L(\gamma, \rho|Z) = \text{constant} + \frac{T}{2}\log(|I - \rho\Phi|) - \frac{TN}{2}\log(\text{tr}[(I - \rho\Phi)\hat{U}'\hat{U}]). \quad (4.5)$$

For a given choice of model complexity  $K$ , maximum likelihood estimation of  $\rho$  and  $\gamma$  can be accomplished by a two-stage optimization method. Given a value of  $\rho$ , an estimate of  $\gamma$  may be found using a standard nonlinear least squares technique (Dennis (1977)) to minimize the last term in (4.5). We then choose the optimal  $\hat{\rho}$  using a simple (one-dimensional) grid-search algorithm.

A key choice to be made in the fitting of such models is that of  $K$ , the number of components. In applications, we may be guided in our choice of  $K$  by a criterion that seeks to balance fidelity to the data with model complexity, such as a modified Akaike Information Criterion (AIC):

$$\text{AIC}_K = -2\ln(\text{maximized likelihood}) + 2p,$$

Here  $p = K * N + K - q$  is the effective number of parameters with  $q = \#\{(j, k) : \beta_{ij} = 0\}$ . Determining the penalty for effective model complexity in this way agrees with the approach taken by Efron et al. (2004) and Zou, Hastie, and Tibshirani (2007), who studied the issue of penalized regression and variable selection.

The algorithm for fitting voxelwise data can be summarized as follows.

- For each  $K = 1, \dots, B$ 
  - For each choice of  $\rho = (0, 0.01, 0.02, 0.03, \dots, 0.99)$ :
    - \* Maximize  $\log L(\gamma, \rho|Z)$  over all choices of  $\gamma$  using a nonlinear least squares technique:
      - For each choice of  $\gamma$ , estimate  $\{\beta_{jk}\}$  using nonnegative least squares fitting.
      - Compute  $\log(L(\gamma, \rho|Z))$  using current values of  $\gamma$  and  $\{\beta_{jk}\}$ .
    - Choose the value of  $\rho$  that maximizes the profile log-likelihood,  $\log(L(\gamma, \rho))$ .
    - Calculate  $AIC_K$  based on the profile log-likelihood.
- Choose the value  $K$  that minimizes  $AIC_K$ .
- Set final estimates of  $\rho$ ,  $\gamma$  and  $\{\beta_{jk}\}$  to be those that maximize the log-likelihood for the chosen  $K$ .

If equal variability among voxels is assumed, then all  $\nu_j$  would be set to one. Otherwise (particularly important in ROI analyses), the fitting algorithm would include estimating the  $\nu_j$ 's iteratively based on

$$\hat{\nu}_j \propto \sum_{i=1}^T w(t_i) \left( y_j(t_i) - \sum_{k=1}^K \hat{\beta}_{jk} f(t_i|\gamma_k) \right)^2,$$

subject to the constraint  $\prod_{j=1}^N \hat{\nu}_j = 1$ .

Estimators of the outcome measure (2.2) can be obtained for each voxel as

$$\hat{V}_j = \sum_{k=1}^K \frac{\hat{\beta}_{jk}}{\hat{\gamma}_k}, \quad j = 1, \dots, N.$$

It is important to emphasize that, although the mixture approach fits all of the voxels or ROI's using the same basis functions, the method does not require that all voxels or ROI's have the same compartmental configuration. For some  $j$  and  $k$ , the estimated coefficients  $\hat{\beta}_{jk}$  may be zero.

A ROI analysis is based on the assumption that the signal intensities within each anatomical region are similar, and therefore the modeling is performed on the mean of all voxelwise data within each ROI across time. Although there might be a spatial contiguity between two ROIs, most of the voxels between two ROIs are not close and therefore the spatial correlation between concentrations measurements made between any two ROIs is weak. To model all ROIs in the

brain, one would apply (4.2) with  $y_j(t_i)$  representing the average signal intensity of the  $j$ th ROI at time  $t_i$  and  $\rho_{jk} = 0$  if  $j \neq k$ . Therefore, for a given  $K$ , the number of the basis functions, the algorithm for fitting the data is simplified to a one-stage procedure as there is no need to estimate  $\rho$ .

## 5. Simulation Studies and Application to Data

In this section we describe the results of application of the proposed modeling approach to both simulated and real datasets. Parameters for the simulation study were set to match those from a data application in a study of the serotonin 1A receptor using the [C-11]WAY-100635 tracer. For both the voxel-level and the ROI-level simulations, we generated data at 20 times points (three 20-second frames, three 60-second frames, three 2-minute frames, two 5-minute frames, and nine 10-minute frames). Data were simulated voxel-by-voxel (and ROI-by-ROI) according to a two-tissue kinetic model, with the kinetic parameters determined from fits to data. Thus, it is important to note that in these simulations, the mixture model is not the “correct” model, as the two components in the “true” model vary across voxels/ROIs. Further details of the imaging protocol, used to guide our simulations, are given in Parsey et al. (2006).

### 5.1. ROI-level simulations

In the first simulation study, data were generated from a two-tissue kinetic model separately for each of 74 regions with the “true” kinetic parameters taken to match those estimated for an actual data set. For this simulation, the kinetic parameters (see (2.2)) had a range of 0.0232 to 0.0452 for  $\beta_1$ , 0.0271 to 0.0519 for  $\beta_2$ , 0.0052 to 0.0530 for  $\gamma_1$ , and 0.0229 to 0.4939 for  $\gamma_2$ . Because the regions have widely differing sizes, the variance for each region is chosen to match the data set as well, with  $\sigma$  ranging from 0.0075 to 0.0429. The weights were set as  $w(t_i) = t_i^e - t_i^s$ .

Results are displayed in Figure 3, with the top left-hand panel representing data from one simulated set of ROIs and the other panels showing the results from 100 such datasets. For the ROI data displayed in the upper left panel, the modified AIC criterion chooses the 4-component model. Estimation of  $V$  for the kinetic model is somewhat unstable, especially for the regions with high binding, for the region-by-region kinetic model. (Even though this is the “correct” model, fitting a two tissue model is often problematic, and thus in practice some constraints are imposed in order to reduce the number of number of free parameters for each region (Parsey et al. (2000)). Spectral analysis tends to overestimate  $V$  for low values of  $V$ . The mixture model, despite being fit to data generated from the “wrong” model, does reasonably well at estimating  $V$ , particularly for the regions with larger  $V$  (which in many applications are the regions of greatest interest).

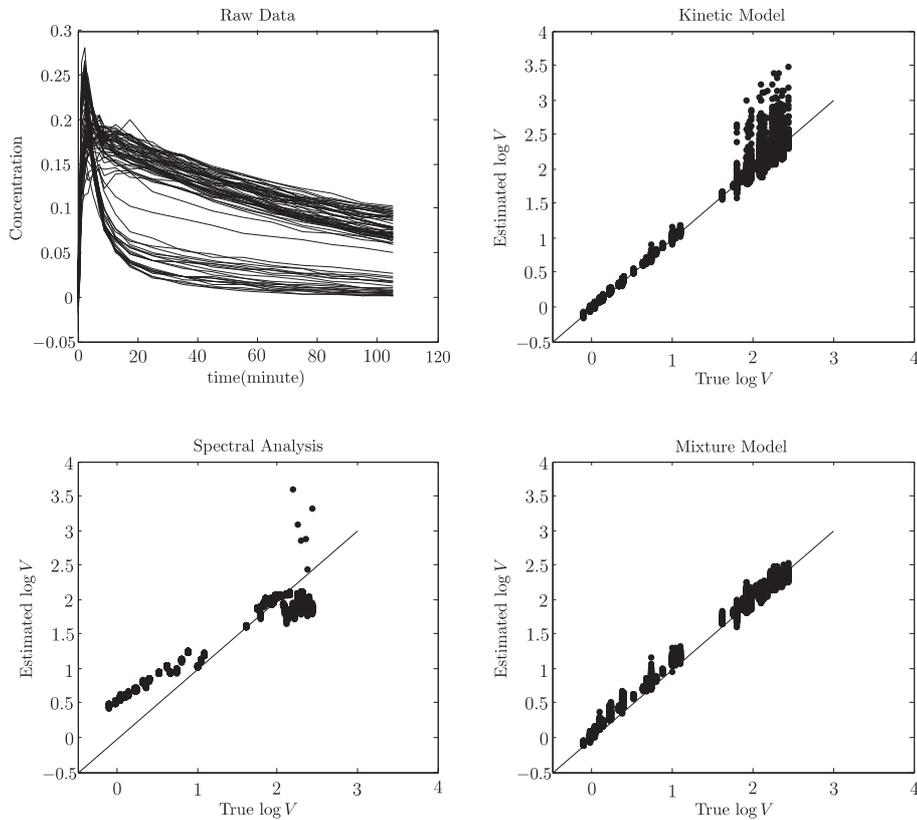


Figure 3. ROI data generated from the two tissue kinetic model separately for each region. Upper left: raw ROI data time-activity curves; upper right: performance of estimated  $V$  using standard two tissue kinetic modeling over 100 simulated data sets; lower left: performance of estimated  $V$  using spectral analysis; lower right: performance of estimated  $V$  using mixture modeling.

## 5.2. Voxel-level simulations

Second, we simulated one transaxial slice of PET data consisting of 6,283 voxels. In agreement with the actual data, the noise level was set to  $\sigma = 0.1$  with the weights set as in Section 5.1. Since the spatial neighborhood effect for voxelwise data is typically quite high, the corresponding parameter,  $\rho$ , was set to 0.9. The “true” kinetic parameters had a range of 0 to 0.0943 for  $\beta_1$ , 0 to 0.4143 for  $\beta_2$ , 0.005 to 0.022 for  $\gamma_1$ , and 0.025 to 1.732 for  $\gamma_2$ .

Results are shown in Figure 4. The plot on the the upper left shows the simulated data for minute 55. The AIC plot indicates a four-component mixture model, and the corresponding  $\hat{V}$  values tend to be quite close to their corresponding true  $V$  values. The mixture approach, choosing model order based on

AIC, provides approximately unbiased estimates. Compared with the commonly used spectral analysis, the mixture modeling gives smaller bias and a smaller standard error in terms of estimating  $\hat{V}$ . The two panels at the bottom of Figure 4 show the four estimated basis functions and the image of estimated  $V$  based on the mixture model. Although the estimated TAC of each voxel is a linear combination of four component functions, most voxels have an estimated coefficient of 0 for at least one component function. For this simulated data set, the average number of non-zero coefficients per voxel is 2.7. The estimated spatial neighborhood effect was  $\hat{\rho} = 0.91$ , quite close to the true value  $\rho = 0.9$ .

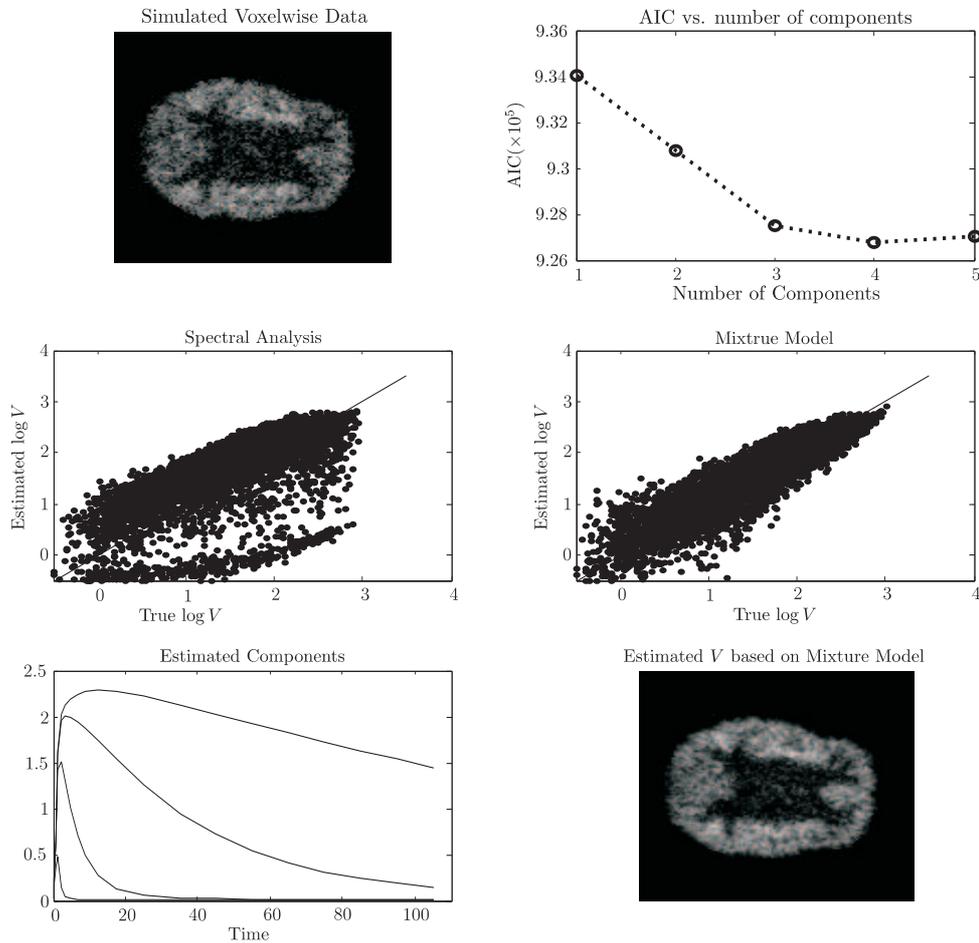


Figure 4. Results from the voxel-level simulation study. Upper left: raw simulated data at time 55; upper right: AIC as a function of number of components; center left: True  $V$  vs.  $\hat{V}$  using spectral analysis; center right: true  $V$  vs.  $\hat{V}$  using mixture modeling; lower left: estimated components of the mixture model; lower right: image of  $\hat{V}$  using mixture modeling.

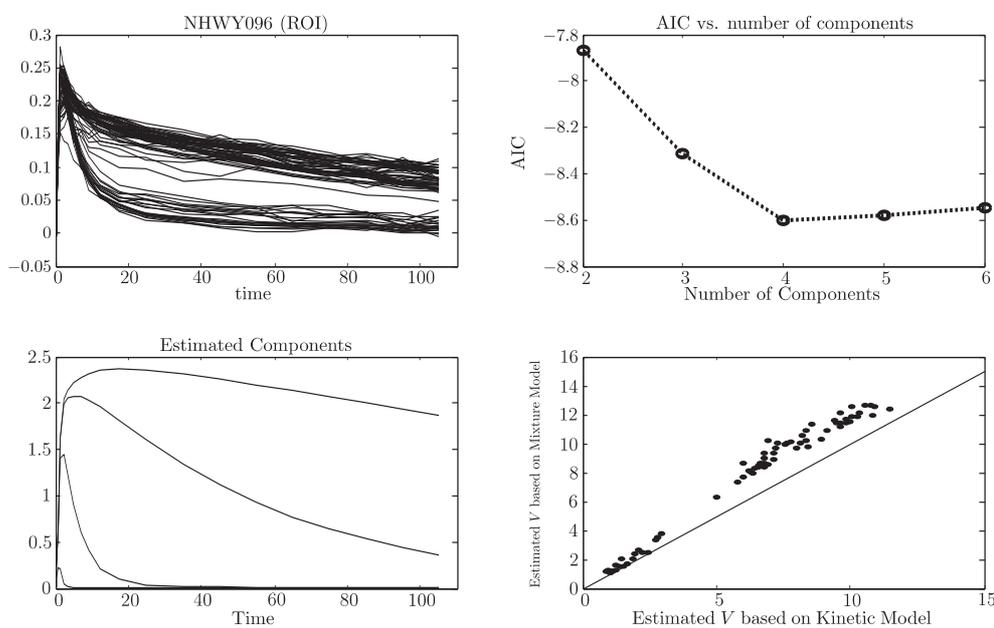


Figure 5. Upper left: TAC data for all ROIs; upper right: AIC as a function of number of components; lower left: the four components estimated by the mixture model; lower right: estimated  $V$  based on traditional two tissue kinetic model vs. estimated  $V$  based on a four-component mixture model.

### 5.3. Application to data

To illustrate the methodology applied to data in which the “truth” is not known, we fit the proposed mixture model to data from one male control subject with the tracer  $[C-11]WAY-100635$ , part of the study reported in Parsey et al. (2006). There are 74 ROIs available, and we applied both a two-tissue kinetic model and the mixture model to each.

The data and model fitting are summarized in Figure 5. As with the simulated data, the AIC plot indicated that a four-component mixture model is best. The mixture model tended to give a higher estimate of  $V$  as compared to the traditional two-component kinetic model.

## 6. Discussion

The methodology described in this paper offers a useful alternative to the usual approach of estimation in PET studies, in which voxels or ROIs are fit one at a time. By allowing more components but requiring constant  $\gamma_k$  values across regions/voxels, the model is flexible but estimation of parameters is quite stable.

Our proposed methodology is related to basis function alternatives to kinetic modeling (Cunningham and Jones (1993) and Gunn et al. (2002)), in which a set

of basis functions of the form (4.1) with a range of  $\gamma_k$  values is created and some constraints are used to ensure sparsity of the solution. Our approach is to use fewer component functions of the same form, but requires that they be constant for all voxels. Compared with our model, that of O'Sullivan (2006) is considerably more general, as he estimates the mixture components nonparametrically. In our particular application, both flexibility and stability are realized by combining these approaches, applying the mixture modeling strategy of O'Sullivan (2006), but with parametrically specified component functions.

One useful way to compare models is in terms of the number of parameters required to fit a dataset, with preference given to more parsimonious models. In application of mixture modeling,  $K$  parameters determining the shape of each component must be determined, and then  $K$  coefficients are estimated for each voxel. Relative model parsimony is achieved because many of the (voxel-specific) coefficients are estimated as zero. In kinetic modeling, on the other hand, the number of parameters to be estimated depends entirely on complexity of the kinetic model selected, e.g., a two-tissue model requires two shape parameters and two coefficients to be estimated for each voxel. In the simulation described in Section 5.2, the total number of non-zero coefficients averaged 2.7 per voxel for the mixture model. A one-tissue kinetic model would require exactly two parameters per voxel; a two-tissue model would require exactly four. Noting the difference in the types of parameters that must be estimated in each of these strategies we observe that, in terms of model parsimony, mixture modeling represents a savings as compared to the two-tissue model. Our primary motivation for the mixture modeling approach is our belief that many of the shape parameters that must be estimated for each voxel in a kinetic modeling analysis are essentially equal, and thus by combining all voxel data together and incorporating this assumption into the model, more stable estimates of these common parameters may be obtained, and that it can be done with fewer parameters. Basis function methods (see, e.g., Gunn et al. (2002)) that automatically estimate model complexity do so in an approximately unbiased way, but since these methods are applied separately for each voxel, they may not achieve similar gains in parsimony relative to standard kinetic modeling.

Computationally, the fitting of the mixture model is reasonably fast, certainly competitive with other methods requiring nonlinear optimization. For a data set consisting of 229,005 voxels and 20 time points, it took about 30 minutes on a Power Mac G5 (dual 2.5GHz) computer. On the same machine, it took less than a minute for each complete (one subject) ROI-level analysis in the simulations.

For application to voxel data, computational savings may be realized by not requiring that all voxels be used to estimate the (common)  $\gamma_k$  values, rather that

a relatively small sample of voxels be taken and the mixture model fit to these for such estimation. Once the component functions are determined, the coefficient parameters  $\beta_{j,k}$  may be estimated very quickly for all voxels. Our experience suggests that the estimation is stable with as few as 5% voxels selected at random (in a study of about 200,000 voxels).

Fitting all voxels at once, as we propose to do, and computing a likelihood function necessitates specifying a spatial model for the original PET data. Spatial relationships among test statistics have been taken into account in the context of controlling the familywise error rate for quite some time (see, e.g., Friston et al. (1991)). In estimation of kinetic parameters in neuroreceptor mapping, fitting is typically done separately for each voxel and thus binding parameters can be estimated without explicitly modeling the spatial covariance of the data. We find that the CAR model described in Section 4 provides a flexible spatial model that is appropriate for PET data. Others have taken differing approaches to modeling spatial correlation of brain imaging data in various contexts (see, e.g., Bowman (2007), Katanoda, Matsuda and Sugishita (2002), Turkheimer et al. (2003) and Maitra and O'Sullivan (1998), among others).

Standard errors of outcome measures may be computed by using an asymptotic approach or a bootstrap method (Ogden and Tarpey (2006)). The bootstrap method requires more computation.

## References

- Bowman, F. D. (2007). Spatiotemporal models for region of interest analyses of functional neuroimaging data. *J. Amer. Statist. Assoc.* **102**, 442-453.
- Chen, K., Huang, S. C. and Yu, D. C. (1991). The effect of measurement errors in plasma radioactivity curve on parameter estimation in positron emission tomography. *Physics in Medicine and Biology* **36**, 1183-1200.
- Cobelli, C., Foster, D. and Toffolo, G. (2000). *Tracer Kinetics in Biomedical Research: From Data to Model*. Springer, New York.
- Cunningham, V. J. and Jones, T. (1993). Spectral analysis of dynamic PET studies. *J. Cerebral Blood Flow and Metabolism* **13**, 15-23.
- Dennis, J. J. (1977). Nonlinear least squares. In *State of the Art in Numerical Analysis* (Edited by D. Jacobs), 269-312. Academic Press, New York.
- Efron, B., Hastie, T., Johnstone, I. and Tibshirani, R. (2004). Least angle regression. *Ann. Statist.* **32**, 407-499.
- Friston, K. J., Frith, C. D., Liddle, P. F. and Frackowiak, R. S. J. (1991). Comparing functional PET images: The assessment of significant change. *J. Cerebral Blood Flow and Metabolism* **11**, 690-699.
- Gunn, R. N., Gunn, S. R. and Cunningham, V. J. (2001). Positron emission tomography compartmental models. *J. Cerebral Blood Flow and Metabolism* **21**, 635-652.
- Gunn, R. N., Gunn, S. R., Turkheimer, F. E., Aston, J. A. D. and Cunningham, V. J. (2002). Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. *J. Cerebral Blood Flow and Metabolism* **22**, 1425-1439.

- Haining, R. (1990). *Spatial Data Analysis in the Social and Environmental Sciences*. Cambridge University Press, Cambridge.
- Katanoda, K., Matsuda, Y. and Sugishita, M. (2002). A spatio-temporal regression model for the analysis of functional MRI data. *Neuroimage* **17**, 1415-1428.
- Lawson, C. L. and Hanson, R. J. (1974). *Solving Least Squares Problems*. Prentice-Hall.
- Maitra, R. and O'Sullivan, F. (1998). Variability assessment in PET and related generalized deconvolution problems. *J. Amer. Statist. Assoc.* **93**, 1340-1356.
- Mintun, M. A., Raichle, M. E., Kilbourn, M. R., Wooten, G. and Welch, M. J. (1984). A quantitative model for the in vivo assessment of drug binding sites with positron emission tomography. *Ann. Neurology* **15**, 217-227.
- Ogden, R. T. and Tarpey, T. (2006). Estimation in regression models with externally estimated parameters. *Biostatistics* **7**, 115-129.
- O'Sullivan, F. (2006). Locally constrained mixture representation of dynamic imaging data from PET and MR studies. *Biostatistics* **7**, 318-338.
- Parsey, R. V., Oquendo, M. A., Ogden, R. T., Olvet, D. M., Simpson, N., Huang, Y.-Y., Van Heertum, R. L., Arango, V. and Mann, J. J. (2006). Altered serotonin 1A binding in major depression: A [carbonyl-C-11]WAY100635 positron emission tomography study. *Biological Psychiatry* **59**, 106-113.
- Parsey, R. V., Slifstein, M., Hwang, D.-R., Abi-Dargham, A., Simpson, N., Mawlawi, O., Guo, N.-N., Van Heertum, R., Mann, J. J. and Laruelle, M. (2000). Validation and reproducibility of measurement of 5-HT1A receptor parameters with [carbonyl-11C]WAY-100635 in humans: Comparison of arterial and reference tissue input functions. *J. Cerebral Blood Flow and Metabolism* **20**, 1111-1133.
- Rencher, A. C. (2002). *Methods of Multivariate Analysis*, Second edition. Wiley, New York.
- Turkheimer, F. E., Aston, J. A. D., Banati, R. B., Riddell, C. and Cunningham, V. J. (2003). A linear wavelet filter for parametric imaging with dynamic PET. *IEEE Trans. Medical Imaging* **22**, 289-301.
- Zou, H., Hastie, T. and Tibshirani, R. (2007). On the "degrees of freedom" of the Lasso. *Ann. Statist.* **35**, 2173-2192.

Biostatistics Division, New York State Psychiatric Institute, New York, NY 10032, U.S.A.

E-mail: [jianghu@pi.cpmc.columbia.edu](mailto:jianghu@pi.cpmc.columbia.edu)

Department of Biostatistics, Columbia University, New York, NY 10032, U.S.A.

E-mail: [to166@columbia.edu](mailto:to166@columbia.edu)

(Received April 2007; accepted May 2008)