

# Spatiotemporal Reconstruction of List-Mode PET Data

Thomas E. Nichols, Jinyi Qi, *Member, IEEE*, Evren Asma, *Student Member, IEEE*, and Richard M. Leahy\*, *Senior Member, IEEE*

**Abstract**—We describe a method for computing a continuous time estimate of tracer density using list-mode positron emission tomography data. The rate function in each voxel is modeled as an inhomogeneous Poisson process whose rate function can be represented using a cubic B-spline basis. The rate functions are estimated by maximizing the likelihood of the arrival times of detected photon pairs over the control vertices of the spline, modified by quadratic spatial and temporal smoothness penalties and a penalty term to enforce nonnegativity. Randoms rate functions are estimated by assuming independence between the spatial and temporal randoms distributions. Similarly, scatter rate functions are estimated by assuming spatiotemporal independence and that the temporal distribution of the scatter is proportional to the temporal distribution of the trues. A quantitative evaluation was performed using simulated data and the method is also demonstrated in a human study using  $^{11}\text{C}$ -raclopride.

**Index Terms**—Calibration, conjugate gradient methods, estimation, Poisson processes, smoothing methods, spline functions.

## I. INTRODUCTION

**D**YNAMIC positron emission tomography (PET) imaging usually involves a sequence of contiguous acquisitions each of which can range in duration from 10 s to over 20 min. Data from each of the frames is independently reconstructed to form a set of images which can be visualized and used to estimate physiological parameters [1]. This approach involves selection of the set of acquisition times, where one must choose between collecting longer scans with good counting statistics but poor temporal resolution, or shorter scans that are noisy but preserve temporal resolution.

List-mode data acquisitions provide extremely high temporal resolution with full spatial resolution. List-mode data can be binned into sinograms, allowing frame durations to be determined after acquisition. Alternatively, the problem of temporal binning can be avoided entirely by directly using the arrival times in the list-mode data to estimate a dynamic image. Such an approach is the focus of this paper.

Manuscript received February 19, 2001; revised March 2, 2002. This work was supported by the National Cancer Institute under Grant R01 CA59794. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was D. Townsend. *Asterisk indicates corresponding author.*

T. E. Nichols is with the Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109 USA.

J. Qi is with the Center for Functional Imaging, Lawrence Berkeley National Laboratory, University of California, Berkeley, CA 94720 USA.

E. Asma and R. M. Leahy are with the Signal and Image Processing Institute, University of Southern California, Los Angeles CA 90089-2564 USA.

\*R. M. Leahy, 3740 McClintock Ave, EEB400, University of Southern California, Los Angeles CA 90089-2564 USA (e-mail: leahy@sipi.usc.edu).

Publisher Item Identifier S 0278-0062(02)04687-6.

Snyder [2] developed a list-mode expectation-maximization (EM) maximum-likelihood (ML) method for estimation of dynamic PET images using inhomogeneous Poisson processes. Each voxel has an associated time-varying tracer density that is modeled using basis functions that are based on assumptions about the physiological processes generating the data, e.g., blood activity curves convolved with a basis of exponentials. The observed list-mode PET data are modeled as inhomogeneous Poisson processes whose rate functions are linear combinations of the dynamic voxel tracer densities. Here, we follow a similar approach but instead work with rate functions formed as a linear combination of known basis functions estimated with a conjugate gradient penalized ML approach. Not only does the linearity of the model lend itself to efficient computation of the estimates, but also we can better represent the dynamic activity seen in experimental data that is not well modeled by more restrictive physiological models. We demonstrate this greater flexibility in Section VI-B, where we discover nonmonotonic behavior in  $^{11}\text{C}$ -raclopride data that is not accounted for in a physiological model.

A second advantage of using list-mode data arises in cases where the number of detected photon pairs in a particular study is far less than the total number of detector pairs. This is often the case in modern three-dimensional (3-D) PET systems which can have in excess of  $10^8$  sinogram elements in a single frame. To reduce this number to manageable proportions, the data are often rebinned by aggregating nearby elements. Alternatively, the raw list-mode data can be stored and the need for rebinning is avoided. Barrett *et al.* [3], [4] describe a list-mode ML method for estimation of a temporally stationary image. While this method will often reduce storage costs and avoid the need for rebinning, the random spatial ordering of the detected events in the list-mode data does not lend itself to fast forward and backprojection and exploitation of the many symmetries in 3-D projection matrices [5], [6]. To avoid this problem we use a hybrid of the sinogram and list-mode formats that allows the reconstruction algorithm to exploit the same matrix symmetries used in our static imaging work [6]. All events in a dynamic study are collected into a single standard sinogram augmented by a “timogram” that contains the arrival times of each event stored so that they are indexed using the values in the associated sinogram.

In this paper we present a method for reconstructing a continuous time estimate of a dynamic PET image using list-mode data and the theory of inhomogeneous Poisson processes. A general B-spline model represents the dynamic activity in each voxel so that the dynamic image is parameterized by a sequence

of control vertex “images” where the control vertices are the coefficients for the spline basis. Tomographic projections of these control vertices produce the control vertices for the rate functions of the inhomogeneous Poisson processes representing coincidence detections between each detector pair. An ML estimate of the control vertices for each voxel can then be computed using the standard likelihood function for inhomogeneous Poisson processes [2], [7]. The final result is a temporally continuous representation of the PET image that utilizes the temporal resolution of list-mode data. Preliminary investigations of this approach are described in our earlier conference articles [8], [9].

Our parameterization of the inhomogeneous Poisson rate function is applicable to any linear combination of basis functions. This form encompasses the parametric imaging work of Matthews [10], Zibulevsky [11], and Snyder [2] and the mixture models of O’Sullivan [12]. We also note that Ollinger [13] used list-mode data to reconstruct rate functions as histograms with adaptive bin sizes; our work could be viewed as a continuous-time extension of this. For this paper we consider only cubic B-splines. The key advantage to B-splines are that they have systematic compact support. In particular, for any point on a cubic spline only four basis functions are nonzero. Also, simple closed forms exist for all derivatives and integrals of a polynomial spline.

Since inhomogeneous Poisson rate functions are unnormalized densities, we note that the density estimation literature using splines is closely related to our work (e.g., [14] and [15]). The standard methods involve exponentiated splines or squared splines. While these implicitly constrain the rate function to be positive, they cannot be represented with a linear basis. As there are substantial computational savings to having a common linear basis for all voxels and projections, we did not pursue these approaches.

The paper is organized as follows. We describe the model and ML method in Sections II and III, respectively. Methods for selecting the spline knot points and methods for randoms and scatter correction are included in Section IV. Computational considerations including re-sorting data into a timogram format and the details of the algorithm used for computing the ML estimate are given in Section V. In Section VI, we demonstrate the performance of the method with quantitative simulations studies and *in vivo* human data.

## II. DYNAMIC MODELING USING INHOMOGENEOUS POISSON PROCESSES

We model the positron emissions from each voxel in the volume as an inhomogeneous Poisson process. The rate function for the voxel represents, to within a scalar calibration factor, the time varying PET tracer density. We parameterize the rate functions using a cubic B-spline basis

$$\eta_j(t) = \sum_{\ell} w_{j\ell} B_{\ell}(t), \quad \eta_j(t) \geq 0 \forall t$$

where  $\eta_j(\cdot)$  is the rate function for voxel  $j$ ,  $w_{j\ell}$  is the  $\ell$ th basis weight (control vertex) for voxel  $j$ , and  $B_{\ell}(t)$  is the  $\ell$ th spline

basis function. The problem of reconstructing the dynamic PET image is then reduced to estimating the control vertices for each voxel.

We denote by  $p_{ij}$  the probability of detecting at detector pair  $i$  a photon pair produced by emission of a positron from voxel  $j$ . The probabilities  $p_{ij}$  are identical to those used in static PET imaging. Here we use the factored matrix forms developed in [6]. Assuming that the detection probabilities are independent and time invariant, it follows that coincidence detection at detector pair  $i$  is also an inhomogeneous Poisson process with rate function

$$\lambda_i(t) = \sum_j p_{ij} \sum_{\ell} w_{j\ell} B_{\ell}(t) = \sum_{\ell} \left( \sum_j p_{ij} w_{j\ell} \right) B_{\ell}(t) \quad (1)$$

where the right-most term demonstrates that the rate functions for the data are also B-splines.

The Poisson process observed at the detectors is corrupted by random and scatter components that can also be modeled as inhomogeneous Poisson processes. Combining the three components, we have the model

$$\lambda_i^*(t) = \lambda_i(t) + r_i(t) + s_i(t)$$

where  $r_i(\cdot)$  and  $s_i(\cdot)$  are the randoms and scatter rate functions for detector pair  $i$  and  $\lambda_i^*(t)$  is the rate function for the process actually observed at detector pair  $i$ . In estimating the rate function parameters  $w_{j\ell}$ , we will assume that the rate functions for the random and scatter components have been determined through a calibration procedure and can be treated as known processes.

For a Poisson process with rate function  $\lambda(t)$ , with  $N$  events observed from time  $T_0$  to  $T_1$  and event arrival times  $a_1, \dots, a_k, \dots, a_N$ , the likelihood function [7] is

$$P(a_1, \dots, a_k, \dots, a_N | \lambda(t)) = \left( \prod_{k=1}^N \lambda(a_k) \right) \cdot \exp \left\{ - \int_{T_0}^{T_1} \lambda(u) du \right\}. \quad (2)$$

For  $N = 0$ , the product is defined as unity.

For the set of independent events recorded in the list-mode data the log likelihood is given by

$$L(\mathcal{D} | \mathcal{W}) = \sum_i \sum_k \log \lambda_i^*(a_{ik}) - \sum_i \int \lambda_i^*(u) du$$

s.t.  $\lambda_i^*(t) \geq 0 \forall t.$  (3)

where  $\mathcal{D}$  denotes the list-mode data and  $\mathcal{W}$  the set of parameters for the rate functions. We represent the data as  $\mathcal{D} = \{\mathbf{x}, \mathbf{a}_1, \dots, \mathbf{a}_i, \dots, \mathbf{a}_I\}$ , where  $\mathbf{x} = (x_1, \dots, x_i, \dots, x_I)$  are the sinogram count data and  $\mathbf{a}_i = (a_{i1}, \dots, a_{ik}, \dots, a_{ix_i})$ , the  $x_i$  event arrival times at detector pair  $i$ . For the B-spline basis,  $\mathcal{W} = \{w_{j\ell} | \ell = 1, \dots, L, j = 1, \dots, J\}$  are the set of basis coefficients. While  $\mathbf{x}$  is a function of  $\mathbf{a}$  and, hence, redundant, we use the sinogram counts to index the arrival times, as described in Section V-A.

### III. PENALIZED ML ESTIMATION

We estimate the image control vertex values that define our dynamic image using penalized ML. The objective function of the statistical model is modified with three regularizing terms

$$L^*(\mathcal{D}|\mathcal{W}) = L(\mathcal{W}|\mathcal{D}) - \alpha\rho(\mathcal{W}) - \beta\phi(\mathcal{W}) - \gamma\nu(\mathcal{W}). \quad (4)$$

The terms  $\rho(\mathcal{W})$  and  $\phi(\mathcal{W})$  regularize temporal and spatial roughness, respectively;  $\nu(\mathcal{W})$  penalizes negativity of the image rate functions;  $\alpha$ ,  $\beta$ , and  $\gamma$  are the tuning parameters. We now describe each of these terms.

We employ a temporal smoothing term to control the roughness of the spline rate functions [16]. The form of the roughness penalty is the integrated squared curvature. For voxel  $j$  this is

$$\int \left\{ \frac{\partial^2}{\partial u^2} \eta_j(u) \right\}^2 du.$$

Fortunately, for cubic splines this quantity has a simple expression, a quadratic form of the control vertices [16, pg. 238]. We denote the symmetric, banded matrix of this quadratic form  $Q$ . Thus, the temporal roughness penalty is given by

$$\rho(\mathcal{W}) = \sum_j \sum_{\ell_1} \sum_{\ell_2} w_{j\ell_1} Q_{\ell_1\ell_2} w_{j\ell_2}.$$

We regularize the estimates of the control vertices using a spatial smoothing function equivalent to the pair-wise quadratic penalty used previously in penalized ML [17] and Bayesian estimation [6] of static PET images

$$\phi(\mathcal{W}) = \sum_{\ell} \sum_j \sum_{j' \in \mathcal{N}_j, j' > j} \kappa_{jj'} (w_{j\ell} - w_{j'\ell})^2.$$

where  $\mathcal{N}_j$  denotes a set of neighbors of voxel  $j$  and  $\kappa_{jj'}$  is the reciprocal of the Euclidean distance between voxel  $j$  and  $j'$ . Other possible choices of the penalty function include the discrete approximation of the thin plate spline bending energy [18] or a nonquadratic edge preserving function such as that described in [19]. In this two-dimensional (2-D) work, we use a second-order neighborhood, where each voxel has eight neighbors.

We note that because the spatial smoothness penalty is not linear, our regularization of the spline coefficient images is not equivalent to penalizing the rate function images. This is principally motivated by computational concerns, but is justified by the use of a B-spline basis. The B-spline basis is well conditioned [16], meaning that small changes in the control vertices produce small changes in the spline function. Hence, if we want two rate functions to be similar, then it is sufficient to constrain their control vertices to be similar.

The optimization method must account for the nonnegativity of the image rate functions  $\eta_j(t)$ . We use unconstrained optimization with a penalty function [20]. The problem is complicated somewhat in that the control vertices themselves are not necessarily nonnegative; instead we need to ensure that the corresponding spline does not become negative. The local extrema of a cubic spline have a closed form, so we initially tried penalizing negative local minima. This approach complicated the gradient and Hessian and made their evaluation prohibitively slow. Instead we simply penalize negative values computed at a finite number of time points. The vector  $\mathbf{z}$  contains the locations at which we enforce positivity. It is constructed by uniformly spacing  $d_z$  points in each interknot interval. Any elements of  $\mathbf{z}$

for which the spline is negative are penalized with the square of the spline value, resulting in the penalty

$$\nu(\mathcal{W}) = \sum_j \sum_m \min \left( 0, \sum_{\ell} w_{j\ell} B_{\ell}(z_m) \right)^2$$

This approach does not necessarily ensure that the spline is nonnegative everywhere. However, we have found that when used in combination with the temporal roughness penalty, the resulting estimates do not become negative, except possibly in the intervals just preceding a large increase in activity.

It is straightforward to show that each of the four terms in the penalized likelihood is concave and, hence, the conjugate gradient method should converge to a global maximum of the cost function.

### IV. CALIBRATION PROCEDURES

#### A. Selection of Knot Spacing

A cubic B-spline basis is defined by knot locations,  $\mathbf{u} = (u_1, \dots, u_{L+4})$ , where  $L \geq 4$  is the number of basis elements and the first and last four knots are identical, to allow discontinuity at the end points. Uniformly spaced knots will not be efficient for most PET tracers since early changes in concentration are much greater than those later in the study. While we do not attempt to adaptively place the knots, in a modest attempt to optimize knot placement, we use the head curve to define knots that produce approximately equal arc lengths, as suggested in [16]. The head curve is a temporal histogram using all of the list-mode data and it serves as an estimate of the average rate function. Once the knot locations are determined, the actual basis functions are computed using recurrence relations as described in [16] and [21].

#### B. Randoms and Scatter Rate Functions

To apply the penalized likelihood estimation procedure described above, we first apply calibration procedures to account for the presence of scattered and random events in the list-mode data. Randoms and scatter correction are essential in extracting accurate quantitative dynamic information from our results. We note that the simple randoms subtraction method that is used in static imaging is not applicable here.

The spatio-temporal randoms distribution is a function of the dynamic tracer distribution. We assume no interaction between the temporal and spatial distribution and scale a fixed spatial estimate over time. While this is a rather crude approximation, it is reasonably accurate due to the very smooth nature of the randoms contribution to the sinogram.

The list-mode data produced by ECAT HR+ and ECAT HR++ PET scanners (CTI Systems, Knoxville, TN) contain both prompt (on-time) and delayed events. Let  $R_i$  denote the total number of delayed events detected at the  $i$ th line of response during the entire acquisition period and  $g(t)$  be an estimate of the dynamic variation in randoms. We compute  $g(t)$  from a spline fit to the delayed-event head-curve,  $g(t) = \sum_{\ell} g_{\ell} B_{\ell}(t)$ ; the control vertices,  $g_{\ell}$ , are obtained as least squares estimates of the head-curve of the delayed events using the same B-spline basis as we use to represent the dynamics of the emission source distribution. We constrain the

least square estimate to be positive and normalize the result to integrate to unity. The estimated randoms rate function is, thus

$$r_i(t) = R_i g(t) = R_i \sum_{\ell} g_{\ell} B_{\ell}(t)$$

We compute the scatter contribution in a similar manner to the randoms, that is we model the scatter rate for each detector pair using an inhomogeneous Poisson process assuming spatiotemporal independence.

The total, time-integrated sinogram is used to estimate the spatial scatter distribution using the simulation method in [22]. Let  $S_i$  denote the estimated scatter contribution at the  $i$ th line of response. Next, we calculate a least-squares spline estimate of the overall true head-curve (prompts minus delays) using the same B-spline basis as we use to represent the dynamics of the emission source distribution. We normalize the result to integrate to unity. Denote this estimate as  $h(t) = \sum_{\ell} h_{\ell} B_{\ell}(t)$  where  $h_{\ell}$  are the control vertices of the true head-curve spline fit. The estimated scatter rate function is then

$$s_i(t) = S_i h(t) = S_i \sum_{\ell} h_{\ell} B_{\ell}(t)$$

## V. COMPUTATIONAL CONSIDERATIONS AND IMAGE ESTIMATION

### A. The “Timogram”

The raw list-mode data is in a form that is inconvenient for computing the gradient of the penalized likelihood function. The list-mode events arrive in random spatial order and, hence, require random rather than sequential access to the projection domain rate functions. We have, therefore, developed a means to store list-mode data in sinogram form while preserving the temporal information. This is achieved using a single standard sinogram, which contains all detected events, augmented by a second file listing the arrival times of all events sorted in projection order. We call this second file the “timogram.” The sinogram is required to indicate how many arrival times to read for each bin. The resulting pair of files can be substantially smaller than either the original list-mode data file or the set of sinograms that would be stored in a conventional dynamic study. We note that Ollinger [13] also resorted list-mode data prior to reconstruction, though his format did not completely eliminate the random spatial order.

ECAT HR+ list-mode data consists of a sequence of 4-byte event words, each either a coincident event or a timing event. The coincident events record the sinogram bin, optional gating information and are identified as “prompt” or “delay.” The timing events are inserted in the list-mode stream every millisecond and they also record time with a 27-bit integer. By re-encoding the arrival time of each coincidence event using 16 bits, we can retain a temporal resolution of 256 ms and a maximum acquisition time of 4.6 hours. Using this format we need only 2 bytes/event in the timogram. Thus, we can discard all of the timing events in the list-mode data and save a factor of two in the space required to store the remaining coincidence arrival times. The space savings from discarding the timing events are significant. For example, in a 90-min scan, the timing events take more space than a 3-D sinogram set and, hence,

the raw list-mode data will always take more space than the sinogram-timogram, even if no coincidences are detected!

The sinogram-timogram format will also be more space efficient than a multiframe sinogram when the space required to store the event arrival times in the timogram is less than the second through  $n$ th sinograms. For example, an 11-frame acquisition is ten frames larger ( $\sim 200$  MB larger) than a sinogram-timogram with no events; only after 200 MB-worth of events, or 100 million counts are stored will the sinogram-timogram be less space efficient.

The sinogram-timogram format could be made even more compact by storing interarrival times and then performing entropy-based compression [23]. The motivation for this is that lines of response with high activity will tend to have short interarrival times, hence, will have many high bits consistently zero, a property that compression can exploit.

### B. Preconditioned Conjugate Gradient (PCG)-Based Reconstruction

A PCG method was used to maximize the objective function. The particular method closely follows our previous work on static reconstructions [6], [24], so we only describe the method briefly here. We use the following preconditioned Polak–Ribiere form of the conjugate gradient method:

$$\begin{aligned} \mathcal{W}^{(n+1)} &= \mathcal{W}^{(n)} + \alpha^{(n)} \mathbf{s}^{(n)} \\ \mathbf{s}^{(n)} &= \mathbf{d}^{(n)} + \beta^{(n-1)} \mathbf{s}^{(n-1)} \\ \mathbf{d}^{(n)} &= \mathbf{C}^{(n)} \mathbf{g}^{(n)} \\ \beta^{(n-1)} &= \frac{(\mathbf{g}^{(n)} - \mathbf{g}^{(n-1)})' \mathbf{d}^{(n)}}{\mathbf{g}^{(n-1)'} \mathbf{d}^{(n-1)}} \end{aligned}$$

where  $\mathbf{g}^{(n)}$  is the gradient vector of the penalized likelihood (4) at  $\mathcal{W} = \mathcal{W}^{(n)}$ ,  $\mathbf{C}^{(n)}$  is a preconditioner, and the step size  $\alpha^{(n)}$  is found using a Newton-Raphson line search; prime (') denotes matrix transpose.

In this study,  $\mathbf{C}^{(n)}$  was chosen analogously to the static PET reconstruction [25] as

$$\mathbf{C}^{(n)} = \text{diag} \left\{ \frac{|w_{j\ell}^{(n)}| + \delta}{\sum_i p_{ij} A_{\ell}} \right\}$$

where  $\delta$  is a small positive number to ensure that  $\mathbf{C}^{(n)}$  is positive definite and  $A_{\ell} = \int B_{\ell}(t) dt$ , the area of each basis element. Here, we set  $\delta$  equal to  $0.01 \max_{j\ell} \{w_{j\ell}^{(n)}\}$ . Note that the standard static preconditioner only has the  $\sum_i p_{ij}$  term in the denominator, which is the (spatial) sensitivity of voxel  $j$ ; we have included  $A_{\ell}$  since it is a measure of the temporal sensitivity of basis element  $\ell$ . While *ad hoc*, we found that this preconditioner had a profound impact on convergence rate (Fig. 1).

The algorithm was initialized with a constant image for which the forward projected rate function matches the average rate of the data after subtracting scatters and randoms. The search vector is initialized by setting  $\mathbf{s}^{(0)} = \mathbf{d}^{(0)}$ . At each iteration we test whether the search vector is an ascent direction, i.e.,  $\mathbf{g}^{(n)'} \mathbf{s}^{(n)} > 0$ . If not, then we reinitialize the PCG algorithm with  $\mathbf{s}^{(n)} = \mathbf{d}^{(n)}$ . To ensure reliable convergence we used a one-sided Newton-Raphson line search: a prospective step size  $\alpha$  was rejected if the sign of the line-search derivative changed.

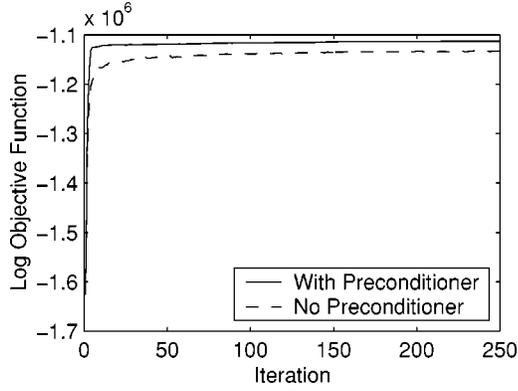


Fig. 1. Convergence of a randomly selected simulation study with and without the use of a preconditioner.

Specifically, the slope of the objective at our current estimate in the direction of the line search is

$$\nabla L^* \left( \mathcal{W}^{(n)} \right)' \mathbf{s}^{(n)} \quad (5)$$

where  $L^*(\cdot)$  is the objective function; the slope in the same direction for a prospective  $\alpha$  is

$$\nabla L^* \left( \mathcal{W}^{(n)} + \alpha \mathbf{s}^{(n)} \right)' \mathbf{s}^{(n)}. \quad (6)$$

If the sign of this second slope disagrees with the original slope, then  $\alpha$  is cut in half.

The logarithm in the likelihood function requires that the line search in (5) is performed with the hard constraint that the forward projected rate function at any arrival time is nonnegative, i.e.,

$$\lambda_i(a_{ik}) \geq 0, \quad \forall i, k.$$

The negativity penalty in (4) is soft, allowing small negative values. The hard constraint can be satisfied by altering the step size in the update step of the PCG algorithm. In essence, we are using a bent, rather than truncated, line search [26]. The bent line search is illustrated in Fig. 2: At a given voxel there are time intervals when the rate function may be essentially zero but the search direction is negative. To prevent this interval from being driven negative we “bend” the search direction, altering it such that it is no longer negative where the rate function is zero. There are many possible alterations that could affect such a change, but we choose the one which causes the smallest  $L^2$  norm change in the direction control vertices.

More precisely, consider a particular voxel  $j$ ; we check each point  $z_m$  to see if

$$\begin{aligned} \eta_j(z_m) &= \sum_{\ell} B_{\ell}(z_m) w_{j\ell} < \epsilon \text{ and} \\ &\sum_{\ell} B_{\ell}(z_m) d_{j\ell} < 0 \end{aligned} \quad (7)$$

where  $d_{j\ell}$  is the  $\ell$ th control vertex of the  $j$ th voxel in  $\mathbf{d}^{(n)}$  and  $\epsilon = 10^{-25}$ . For each  $z_m$  for which this condition holds, we alter the direction as follows

$$d_{j\ell}^{\text{bent}} = d_{j\ell} - \frac{\sum_{\ell'} B_{\ell'}(z_m) d_{j\ell'}}{\sum_{\ell'} (B_{\ell'}(z_m))^2} B_{\ell}(z_m). \quad (8)$$

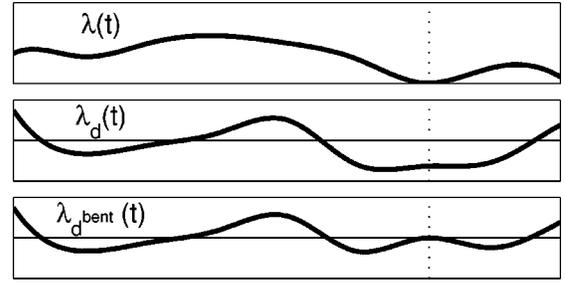


Fig. 2. Illustration of bent line search. If the rate function  $\lambda(t)$  is zero at some point and the search direction  $\lambda_d(t)$  is negative, we “bend” the search direction ( $\lambda_{d^{\text{bent}}}$ ) so that we do not push the rate negative at that point.

This rule can be shown to minimize  $\sum_{\ell} (d_{j\ell}^{\text{bent}} - d_{j\ell})^2$  subject to the constraint  $\sum_{\ell} B_{\ell}(z_m) d_{j\ell}^{\text{bent}} = 0$ . We check each  $z_m$  in order and immediately update the direction each time (7) is satisfied.

This issue of how to choose the penalty parameters  $\alpha$ ,  $\beta$ , and  $\gamma$  is difficult. In static maximum *a posteriori* (MAP) reconstruction, parameters for the smoothing priors can be chosen using ML principles by interpreting the smoothing terms as the Gibbs energy functions of a Markov random field model [27] or can be chosen to achieve a given resolution [28]. These approaches can be extended to include the temporal smoothing parameters, but this is beyond the scope of this paper. We have adjusted the spatial and temporal parameters to obtain resolution typical of that used in clinical PET studies. Objective methods for selecting the weighting of the nonnegativity constraint are needed.

## VI. SIMULATION STUDIES AND PERFORMANCE EVALUATION

We evaluated our method with simulated and real data. We simulated a blood flow data set using a single slice of the Hoffman brain phantom. We evaluated the simulated data on the basis of temporally averaged rate accuracy as described below. We also applied our approach to real data from a  $^{11}\text{C}$ -raclopride study; our subjective evaluations focused on tissues that are known to have distinctly different dynamics with this tracer.

### A. Simulation Study

The simulated data were based on a simplified model of the dynamics of a bolus injection of  $^{15}\text{O}$ -water using tissue time activity curves generated by the Kety autoradiographic model (Fig. 3, cf. [29, Fig. 3B]). We chose two extreme curves, one corresponding to very high blood flow, one to very low blood flow. White matter voxels were assigned to have low blood flow, gray matter voxels to have high blood flow; cerebral spinal fluid was assigned the white matter curve at 20% magnitude. Within the circular support of the reconstruction, the proportion of voxel types were as follows: White matter 15%, gray matter 21%, cerebral spinal fluid 4%, and the remainder background. We used an 11-element B-spline basis with support from 0 s to 140 s; the spacing of the knot locations was determined by dividing a medium blood flow curve into eight equal arc-lengths. We used seven negativity penalty points ( $d_z$ ) in each knot interval. 50 realizations each with approximately 500 000 counts were generated. No scatter or randoms were added or estimated and the

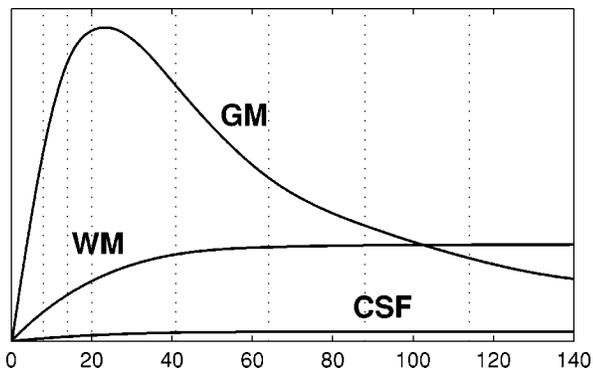


Fig. 3. Three temporal functions defining truth in the simulation study. Gray matter (GM) has highest peak and quick wash out, white matter (WM) has slow rise and cerebral spinal fluid (CSF) has negligible activity. Dotted vertical lines indicate knot locations.

simulations used the same system model as the reconstruction. Each realization was reconstructed using 100 iterations.

We evaluated the method's ability to estimate average rate over time intervals centered about three times,  $t = 23, 45,$  and  $100$  s. As a comparison, we estimated average rate using static MAP reconstructions based on events arriving in the same time intervals; the MAP reconstruction method was based on the same system model and spatial prior and is the base from which the spline reconstruction code was developed. With our method we have two means to adjust temporal resolution: via the temporal regularization parameter  $\alpha$  and by adjusting the length of the time interval; we did both. We considered 1-, 2-, 4-, 10-, and 20-s intervals. We also varied spatial resolution in both our method and the static method by adjusting the  $\beta$  parameter. The parameter ranges in the two different methods were chosen to approximately match noise and resolution properties. Since measuring the spatial resolution of the spline method is not straightforward, we measured the point spread function of the static reconstruction for the 20-s interval about  $t = 23$ ; we found this to be 9.2-mm full-width at half-maximum for the second of three spatial resolutions considered.

We considered voxel-wise squared bias and variance averaged over the whole image and region of interest (ROI) squared bias and variance. Four ROIs were defined based on the left and right caudate and putamen, the two structures comprising the striatum; note that these bilateral regions are not symmetric in the Hoffman phantom. In addition to their neurological relevance, these structures were chosen because they were surrounded by white matter and cerebral spinal fluid and, hence, susceptible to bias from limited spatial resolution.

While variance has a well known unbiased estimator, the natural estimator of squared bias is, itself, biased. With an assumption of Gaussian noise of magnitude  $\sigma^2$ , we show in Appendix A that the expected value of squared bias estimated from 50 Monte Carlo realizations is positively biased by  $\sigma^2/50$ . Since  $\sigma^2$  itself is unknown, correcting this bias with an estimated variance for each voxel or small ROIs is not useful, since the correction will increase variability of the squared bias estimator, even leading to negative estimates! But for image averages and large ROIs, we found this correction to be useful and we have applied it to the results below.

Fig. 4 shows average voxel-wise bias-variance curves for the three time points considered. The left plot shows performance as temporal resolution is varied, the right plot shows performance as spatial resolution is varied. In each plot there are three sets of curves shown, one for each time point (23, 45, and 100 s). Note that for each time point, for a given variance, the spline reconstruction has lower bias.

For the left plot showing the varying temporal resolution, each spline result (solid lines) shows the performance over different temporal penalty parameters, while each static result (dashed lines) shows the performance over different length temporal intervals (1, 2, 4, 10, and 20 s). Note that there are a pair of nearly-overlapping spline results for each time point: While we have a spline result for each of the five temporal interval lengths, we only show those for 1 and 20 s; the 1-s result is above and to the right of the 20-s result. Observe that the static results have nearly constant bias, which is to be expected since the spatial resolution is fixed on this plot. Also note that, with decreasing temporal resolution, the static and spline estimates will eventually converge, as both will essentially estimate a single static image based on all counts.

The right plot shows the results over different spatial resolutions. On this plot, both the spline and static results show the usual bias-variance tradeoffs.

Fig. 5 shows ROI bias-variance curves for  $t = 23$  s, the time of the mode of the high flow curve. The top two plots are for the caudate ROIs, the bottom two are for the putamen ROIs. The left plots show performance as temporal resolution is varied, the right plots show performance as spatial resolution is varied. The plots show a similar performance to the voxelwise averaged results: Varying temporal resolution, both 1- and 20-s interval averaged spline results are superior to the static results, which have near-constant bias; for both spatial and temporal results, for a given variance, the spline reconstruction has lower bias.

## B. Human Studies

The  $^{11}\text{C}$ -raclopride study was performed using data from the ECAT HR + scanner. We used a 15-element B-spline basis with support over the whole acquisition duration of 95 min and knot spacing was determined by equal spacing of 11 points along the head-curve. Single slice rebinning was used to create 95 2-D list-mode data sets that were each independently reconstructed to form the final four-dimensional (4-D) image. Fig. 6 shows one slice of the tracer distribution for the  $^{11}\text{C}$  study after 30 iterations. On a 450-MHz Sun Ultra workstation this took approximately 2 min/iteration/slice, or about one hour/image for 30 iterations. Increased specific binding of  $^{11}\text{C}$ -raclopride in the striatum is seen from 150 to 1200 s in Figs. 6(c) and 6(d). Fig. 6(b) shows three time activity curves for uptake in the scalp, cortex and striatum for the regions of interest marked in Fig. 6(a). Note the faster rise in the sinus, reflecting the early arrival of the tracer in the blood relative to brain tissue. These two observations demonstrate the ability of our method to retain spatial *and* temporal contrast.

We were concerned that the nonmonotonic decrease in tracer density in striatum and cortex after peak uptake (approximately 700 s) was artifactual. We reconstructed this data using different

## Average Rate Estimation – Voxelwise Average

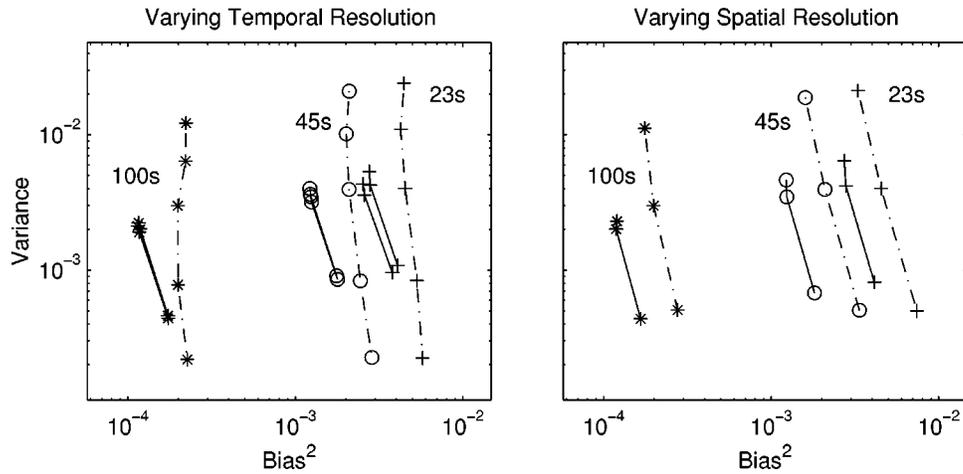
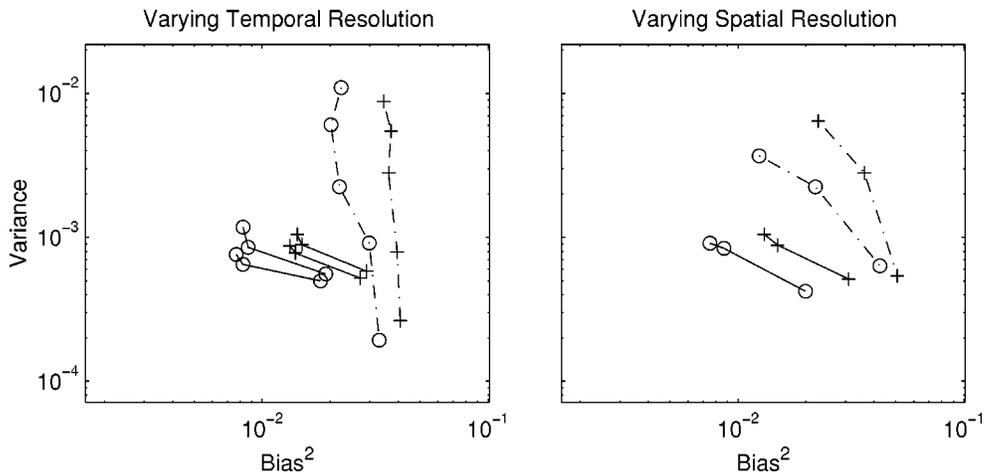


Fig. 4. Voxel-wise average results of average rate estimation simulations, for 23, 45, and 100 s; solid lines are spline estimates, dashed lines are static estimates. For each spatial and temporal resolution, 50 Monte Carlo realizations were created to produce images of squared bias and variance; the average of these images is plotted. There are two ways to vary temporal resolution with the spline estimates and, hence, two curves for each spline result in the left graph, though they closely overlap for the 45- and 100-s results. The lower curve is for averaging over a 20-s interval, the upper curve is for a 1-s interval.

## Average Rate Estimation at 23s – L &amp; R Caudate ROIs



## Average Rate Estimation at 23s – Putamen ROIs

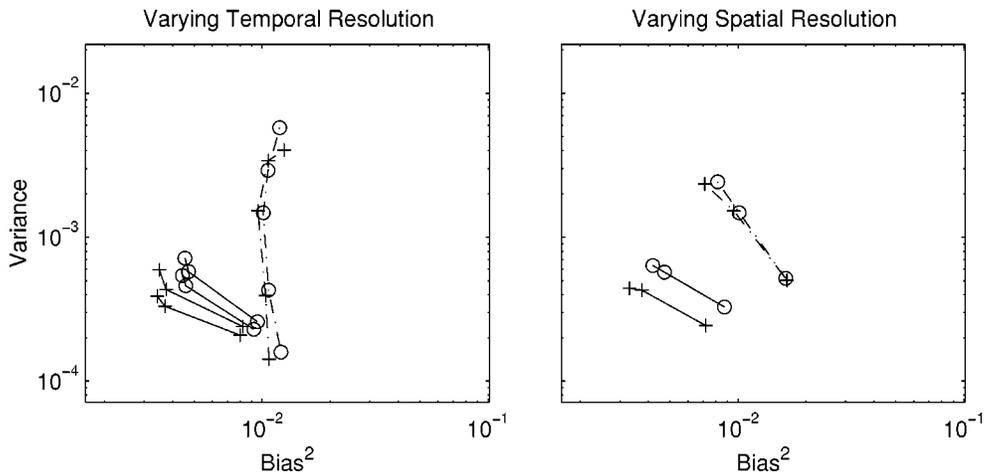


Fig. 5. ROI results of average rate estimation simulations, for time point 23 s; solid lines are spline estimates, dashed lines are static estimates; plus marks (+) indicate left ROIs, circles (o) right ROIs. For each spatial and temporal resolution, ROI squared bias and variance were calculated based on 50 Monte Carlo realizations. Each plot shows the left and right ROIs. There are two curves for each spline result in the left two graphs; the lower curve is for averaging over a 20-s interval, the upper curve is for a 1-s interval.

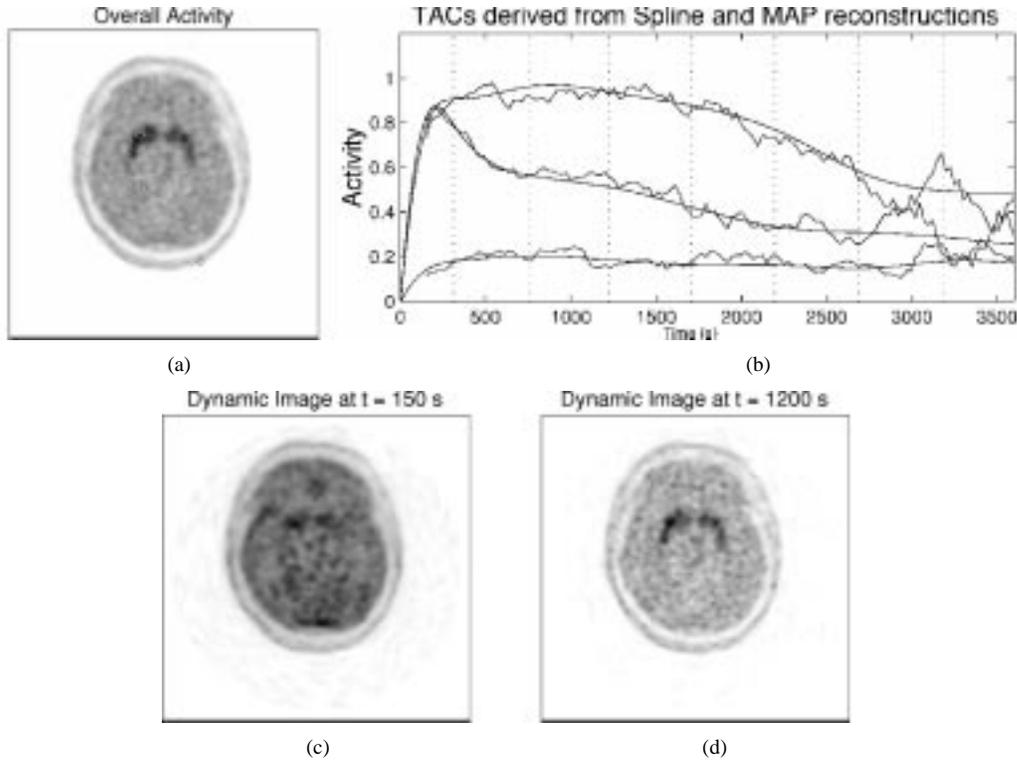


Fig. 6.  $^{11}\text{C}$ -raclopride study from EXACT HR++. (a) A 2-D transaxial section through striatum showing activity integrated over the full 5700-s acquisition. (b) Decay-corrected time activity curves averaged over 25-voxel ROIs for scalp (lower curve), cortex (middle curve), and striatum (upper curve)—the smooth lines are from the spline reconstruction, the jagged lines are from the static reconstructions. Sample images of the continuous time reconstructions obtained by sampling the B-spline curves at each voxel at (c) time  $t = 150$  s and (d)  $t = 1200$  s.

knot locations and different numbers of knots. We found little change in the estimated activity. This appears to indicate that the effect was not due to our curve parameterization. We also binned the data into 300-s static datasets with 25-s shifts (i.e., 275-s overlap) and reconstructed each independently; while the time courses extracted from these images were much noisier, the nonmonotonicity was again evident. This decrease could be due to head motion or an unexpected physiological effect.

## VII. DISCUSSION AND CONCLUSION

We have presented an approach and preliminary results for estimating continuous time dynamic PET images from list-mode PET data. We modeled the dynamic tracer density as an inhomogeneous Poisson process and parameterized the rate functions with a B-spline basis. We introduced the timogram as a means to compactly represent the temporal information of list-mode data. The B-spline basis and the timogram's spatial ordering both contribute to an efficient implementation that makes the creation of continuous time reconstructions feasible.

We have described and implemented approaches to scatter and randoms correction based on spatiotemporal independence, though have left dead time for future work. We have presented basic performance analysis with arbitrarily chosen tuning parameters for spatial and temporal regularization. For the considered range, our method out-performs a comparable static estimate. While our method is 2-D, generalization to three dimensions is straight forward and simply involves use of 3-D system matrices ( $(p_{ij})$ ) instead of the 2-D ones we used here.

We believe that direct estimation of 4-D spatiotemporal tracer distributions based on list-mode is an important direction for PET. Many of the central issues in image reconstruction will have to be revisited (e.g., scatter and randoms) and updated to account for the lack of temporal stationarity. We have introduced only basic solutions for these problems, but have demonstrated that continuous-time image reconstruction is a tractable problem and that it can offer superior performance relative to static methods.

## APPENDIX

### BIAS OF SQUARED BIAS ESTIMATOR

Here, we show that the natural estimator for squared bias is, itself, biased. Let  $f$  be the true rate at a given voxel or ROI. Let  $\hat{f}_k$  be the rate estimate at voxel  $i$  from realization  $k$ ,  $k \in \{1, \dots, K\}$ . Let  $\omega_k$  be the error at voxel  $i$  associated with the  $k$ th realization. We assume that this error is normally distributed, centered at  $b$  with variance  $\sigma^2$ . That is, the bias of the reconstruction method is  $b$  and its variance is  $\sigma^2$

$$\begin{aligned}\hat{f}_k &= f + \omega_k \\ \omega_k &\sim \mathcal{N}(b, \sigma^2).\end{aligned}$$

We are interested in the variance and squared bias. We first consider variance; the unbiased estimator is

$$\hat{\sigma}^2 = \frac{1}{K-1} \sum_k (\hat{f}_k - \hat{f})^2$$

where  $\hat{f}$  is the average of the Monte Carlo realizations, that is  $\hat{f} = (1/K) \sum_k \hat{f}_k$ .

Now consider the squared bias

$$\hat{b}^2 = (\hat{f} - f)^2.$$

To work out the expectation, first note

$$\begin{aligned} \hat{f} - f &= \frac{1}{K} \sum_k (f + \omega_k) - f \\ &= \frac{1}{K} \sum_k \omega_k \\ &= b + \frac{1}{K} \sum_k (\omega_k - b). \end{aligned}$$

Then

$$\begin{aligned} \mathbb{E}[\hat{b}^2] &= \mathbb{E}\left[(\hat{f} - f)^2\right] \\ &= \mathbb{E}\left[\left(b + \frac{1}{K} \sum_k (\omega_k - b)\right)^2\right] \\ &= \mathbb{E}\left[b^2 + 2b \frac{1}{K} \sum_k (\omega_k - b) + \left(\frac{1}{K} \sum_k (\omega_k - b)\right)^2\right] \\ &= b^2 + 0 + \frac{\sigma^2}{K^2} \mathbb{E}\left[\left(\sum_k \frac{\omega_k - b}{\sigma}\right)^2\right] \\ &= b^2 + \frac{\sigma^2}{K}. \end{aligned} \quad (9)$$

The inner term in (9) vanishes because  $\mathbb{E}(\omega_k - b) = 0$ . In (10),  $(\omega_k - b)/\sigma$  is a standard normal, hence, its sum of squares is a  $\chi^2$  random variable with degrees of freedom  $K$  and, hence, has expectation  $K$ .

#### ACKNOWLEDGMENT

This work was conducted while T. E. Nichols was a Visiting Scholar at the University of Southern California, Los Angeles. The authors would like to thank D. Townsend for providing list-mode data and for information on the HR+ list-mode format. They would also like to thank P. Bloomfield for generously sharing his list-mode decoder source code and the  $^{11}\text{C}$ -raclopride data.

#### REFERENCES

- [1] S. Huang and M. Phelps, "Principles of tracer kinetic modeling in positron emission tomography and autoradiography," in *Principles and Applications for the Brain and Heart*. New York: Raven, 1986, Positron Emission Tomography and Autoradiography, ch. 7.
- [2] D. Snyder, "Parameter estimation for dynamic studies in emission-tomography systems having list-mode data," *IEEE Trans. Nucl. Sci.*, vol. 31, pp. 925–931, Apr. 1984.
- [3] H. Barrett, T. White, and L. Parra, "List-mode likelihood," *J. Opt. Soc. Amer.*, vol. 14, no. 11, pp. 2914–2923, 1997.
- [4] L. Parra and H. Barrett, "List-mode likelihood: EM algorithm and image quality estimation demonstrated on 2-D PET," *IEEE Trans. Med. Imag.*, vol. 17, pp. 228–235, Apr. 1998.
- [5] C. Johnson, Y. Yan, R. Carson, R. Martino, and M. Daube-Witherspoon, "A system for the 3-D reconstruction of retracted-septa PET data using the EM algorithm," *IEEE Trans. Nucl. Sci.*, vol. 42, pp. 1223–1227, Aug. 1995.
- [6] J. Qi, R. Leahy, S. Cherry, A. Chatzioannou, and T. Farquhar, "High resolution 3D Bayesian image reconstruction using the microPET small animal scanner," *Phys. Med. Biol.*, vol. 43, no. 4, pp. 1001–1013, 1998.
- [7] D. Snyder and M. Miller, *Random Point Processes in Time and Space*, 2nd ed. New York: Springer-Verlag, 1991.
- [8] T. Nichols, J. Qi, and R. Leahy, "Continuous time dynamic PET imaging using list mode data," in *Lecture Notes in Computer Science*. Berlin, Germany: Springer-Verlag, 1999, vol. 1613, Information Processing in Medical Imaging, pp. 98–111.
- [9] E. Asma, T. E. Nichols, J. Qi, and R. M. Leahy, "4-D PET image reconstruction from list mode data," presented at the IEEE Nuclear Science Symp. Medical Imaging Conf., Lyon, France, 2000.
- [10] J. Matthews, D. Bailey, P. Price, and V. Cunningham, "The direct calculation of parametric images from dynamic PET data using maximum-likelihood iterative reconstruction," *Phys. Med. Biol.*, vol. 42, pp. 1155–1173, 1997.
- [11] M. Zibulevsky, "MI reconstruction of dynamic PET images from projections and clist," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 1, Seattle, WA, 1999, pp. 889–891.
- [12] F. O'Sullivan, "Image radiotracer model parameters in PET: A mixture analysis approach," *IEEE Trans. Med. Imag.*, vol. 12, pp. 399–412, Sept. 1993.
- [13] J. Ollinger, "Algorithms for parameter estimation in dynamic tracer studies using positron emission tomography," Ph.D. dissertation, Washington Univ. Sch. Med., St. Louis, MO, Sept. 1986.
- [14] G. Wahba, "Interpolating spline methods for density estimation. I: Equi-spaced knots," *Ann. Statist.*, vol. 3, pp. 30–48, 1975.
- [15] C. Gu and C. Qiu, "Smoothing spline density estimation: Theory," *Ann. Statist.*, vol. 21, pp. 217–234, 1993.
- [16] C. de Boor, *A Practical Guide to Splines*. New York: Springer-Verlag, 1978, vol. 27, Applied Mathematical Sciences.
- [17] J. Fessler, "Penalized weighted least-squares image reconstruction for PET," *IEEE Trans. Med. Imag.*, vol. 13, pp. 290–300, June 1994.
- [18] S.-J. Lee, A. Rangarajan, and G. Gindi, "Bayesian image reconstruction in SPECT using higher order mechanical models as priors," *IEEE Trans. Med. Imag.*, vol. 14, pp. 669–680, Dec. 1995.
- [19] S. Geman and D. McClure, "Statistical methods for tomographic image reconstruction," *Bull. ISI (Proc. 46th Session ISI)*, vol. 52, 1987.
- [20] D. Luenberger, *Linear and Nonlinear Programming*. Reading, MA: Addison-Wesley, 1989.
- [21] R. Bartels, J. Beatty, and B. Barsky, *An Introduction to Splines for use in Computer Graphics and Geometric Modeling*. Los Altos, CA: Morgan Kaufmann, 1986.
- [22] C. Watson, D. Newport, M. Casey, R. deKemp, R. Beanlands, and M. Schmand, "Evaluation of simulation based scatter correction for 3-D PET cardiac imaging," *IEEE Trans. Nucl. Sci.*, vol. 44, pp. 90–97, Feb. 1997.
- [23] D. Huffman, "A method for the construction of minimum-redundancy codes," in *Proc. Inst. Radio Eng.*, vol. 40, 1952, pp. 1098–1101.
- [24] E. Mumcuoglu, R. Leahy, S. Cherry, and Z. Zhou, "Fast gradient-based methods for Bayesian reconstruction of transmission and emission PET images," *IEEE Trans. Med. Imag.*, vol. 13, pp. 687–701, Dec. 1994.
- [25] J. Qi, R. Leahy, C. Hsu, T. Farquhar, and S. Cherry, "Fully 3-D Bayesian image reconstruction for ECAT EXACT HR+," *IEEE Trans. Nucl. Sci.*, vol. 45, pp. 1096–1103, June 1998.
- [26] L. Kaufman, "Maximum likelihood, least squares and penalized least squares for PET," *IEEE Trans. Med. Imag.*, vol. 12, pp. 200–214, June 1993.
- [27] Z. Zhou, R. Leahy, and J. Qi, "Approximate maximum-likelihood hyper-parameter estimation for gibbs-priors," *IEEE Trans. Image Processing*, pp. 844–861, June 1997.
- [28] J. Qi and R. M. Leahy, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Trans. Med. Imag.*, vol. 19, pp. 493–506, May 2000.
- [29] P. Herscovitch, J. Markham, and M. Raichle, "Brain blood flow measured with intravenous  $\text{H}_2$   $^{15}\text{O}$ . I. theory and error analysis," *J. Nucl. Med.*, vol. 24, pp. 782–789, 1983.