Mean and Covariance Properties of Dynamic PET Reconstructions From List-Mode Data

Evren Asma and Richard M. Leahy*

Abstract—**We derive computationally efficient methods for the estimation of the mean and variance properties of penalized likelihood dynamic positron emission tomography (PET) images. This allows us to predict the accuracy of reconstructed activity estimates and to compare reconstruction algorithms theoretically. We combine a bin-mode approach in which data is modeled as a collection of independent Poisson random variables at each spatiotemporal bin with the space-time separabilities in the imaging equation and penalties to derive rapidly computable analytic mean and variance approximations. We use these approximations to compare bias/variance properties of our dynamic PET image reconstruction algorithm with those of multiframe static PET reconstructions.**

Index Terms—**Dynamic PET, Fisher information matrix, image reconstruction, uniform resolution.**

I. Introduction

MAXIMUM *a posteriori* (MAP) or penalized ML image reconstruction methods combine accurate physical and statistical modeling of the coincidence detection process in positron emission tomography (PET) to produce images with improved resolution and noise performances compared to filtered-backprojection methods. Statistical properties of *single frame* MAP reconstructions as characterized by the mean and variance of the reconstructed images have previously been investigated theoretically both on an algorithm dependent, iteration-by-iteration basis [1]–[3] and also by assuming that the fixed-point of the objective function is exactly achieved [5], [6]. Qi and Leahy derived computationally efficient Fourier-based approximations for the expressions in [5] and validated their approximations on simulated 3-D datasets [8]. Abbey [10] also used the fixed-point approach to evaluate the mean and variance properties of implicitly defined estimators. More recently, Qi united the iteration-based and fixed-point approaches [11] by deriving the fixed-point results from the iteration-based analysis as the number of iterations approaches infinity.

Spatial resolution properties of static PET image reconstruction methods have also been well investigated, often with the objective of achieving uniform and isotropic spatial resolution. Fessler and Rogers [6] replaced the constant smoothing parameter with an image and position dependent smoothing parameter which provided approximately uniform resolution. Qi and Leahy [7] derived computationally efficient methods for computing the local contrast recovery coefficient, which is a measure of resolution, and applied their results to real 3-D PET data. Stayman and Fessler removed the asymmetries in the local impulse response and described an optimal fit of the local impulse response to a predefined target impulse response function [12], [13]. Nuyts and Fessler [18] compared the resolution properties of images reconstructed using penalized-likelihood versus those of images reconstructed with maximum-likelihood and postsmoothed with the penalized-likelihood impulse response. Mustafovic *et al.* [14], [15] showed that uniform spatial resolution could also be achieved by applying a spatially varying filter between the iterations of the (ordered subsets) expectation-maximization or separable paraboloidal surrogates algorithms [4].

In contrast to the growing number of papers on the analysis of single frame penalized ML reconstruction, there is little previous work on mean, variance and spatiotemporal resolution properties of penalized ML dynamic PET images. This is mainly because dynamic imaging is traditionally performed as a series of static reconstructions in which case the results derived for spatial properties of static PET images are applicable to each frame. In such multiframe static reconstructions, temporal properties are determined by the lengths of the frames.

In this paper we investigate the mean and variance properties of the time-activity curves reconstructed with our dynamic penalized ML algorithm, which operates directly on list-mode data [16], [19], through analytic approximations similar to those developed in [5]. We derive computationally efficient approximate methods for the estimation of the mean and variances of dynamic PET images and use these approximations to theoretically compare bias/variance properties of dynamic PET reconstructions with those of multiframe static reconstructions.

II. Theory

A. Objective Functions

The details of the dynamic penalized ML reconstruction method that we analyze below can be found in [16] and are not repeated here. Using the likelihood function of event arrival times in an inhomogeneous Poisson process, we obtain the continuous-time log-likelihood function of the arrival times as

Manuscript received April 26, 2005; revised September 19, 2005. This work was supported in part by the National Institute of Biomedical Imaging and Bioengineering under Grant R01 EB000363. The Associate Editor responsible for coordinating the review of this paper and recommending its publication was J. Fessler. *Asterisk indicates corresponding author.*

E. Asma was with the Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA. He is currently with General Electric Global Research, Functional Imaging Laboratory, Niskayuna, NY 12309 USA.

*R. M. Leahy is with the Signal and Image Processing Institute, University of Southern California, 3740 McClintock Ave., EEB400, Los Angeles, CA 90089-2564 USA (e-mail: leahy@sipi.usc.edu).

Digital Object Identifier 10.1109/TMI.2005.859716

a function of the spatiotemporal basis function coefficients as follows [16]:

$$L_{\mathrm{CT}}(\mathbf{w}) = \sum_{i=1}^{n_d}\sum_{k=1}^{x_i}\log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}B_\ell(a_{ik})\right) - \sum_{i=1}^{n_d}\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_0^T B_\ell(t)\,dt\right) \quad (1)$$

where n_d is the number of detector pairs, n_v is the number of voxels, n_b is the number of temporal basis functions $B_\ell(t), x_i$ is the number of events at detector pair i, a_{ik} is the arrival time of the kth event at detector pair i, p_{ij} is the probability that an event at voxel j is detected at detector pair i, T is the scan duration, $\mathbf{w}$ is the $n_v \times n_b$ vector containing the control vertex vectors of all voxels (i.e., $\mathbf{w} = [\mathbf{w}_1^T \ \mathbf{w}_2^T \ \cdots \ \mathbf{w}_{n_v}^T]^T$ where $\mathbf{w}_j$ is the $n_b \times 1$ control vertex vector for a single voxel) and the subscript "CT" stands for continuous-time. The rate function, $\eta_j(t)$ at the jth voxel is computed from the control vertices as $\eta_j(t) = \sum_{\ell=1}^{n_b} w_{j\ell}B_\ell(t)$.

Computing the Fisher information matrix (FIM) using this log-likelihood function with $-E((\partial^2 L(\mathbf{w}))/(\partial^2 \mathbf{w}))$ leads to an intractable form. The major difficulty arises in the computation of expected values over the event arrival times at each detector pair. Under the inhomogeneous Poisson process model, the dimension of the integral at detector pair i is a Poisson random variable with mean $\int_0^T r_i(t)\,dt$ where $r_i(t)$ denotes the rate function at detector pair i. This requires the computation of an infinite number of multidimensional integrals (because the expectation operation requires a sum over all possible values in the Poisson distribution) and, therefore, significantly complicates a theoretical analysis.

To overcome these difficulties in working with L_{CT}, we divide the scan duration into N time bins and model our observation at each detector pair as the $N \times 1$ vector whose nth element is the number of events observed during time bin n at that detector pair. Under this model, our observation at detector pair i is given by $\mathbf{y}_i = [y_i^{(1)}, y_i^{(2)} \cdots y_i^{(N)}]^T$ where the superscripts denote the temporal bin index. When we concatenate the observation vectors at all detector pairs into a single $n_d \times N$ vector where n_d is the total number of detector pairs, our observation vector becomes $\mathbf{y} = [\mathbf{y}_1^T \ \mathbf{y}_2^T \ \ldots \ \mathbf{y}_{n_d}^T]^T$.

The bin-mode log-likelihood for data binned into such temporal bins is given by

$$L_{\mathrm{BM}}(\mathbf{w}) = \sum_{i=1}^{n_d}\sum_{n=1}^{N} y_i^{(n)}\log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right) - \sum_{i=1}^{n_d}\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_0^T B_\ell(t)\,dt\right). \quad (2)$$

As bin-width approaches zero, the bin-mode log-likelihood function given by (2) and, therefore, its maximizer, approaches the continuous time log-likelihood and its maximizer, respectively (see Appendix). This result allows us to work with bin-mode data and extend the results to the continuous time case by letting the bin width approach zero.

Both continuous-time and bin-mode objective functions are obtained by adding the penalty terms to the respective log-likelihoods

$$\Phi_{\mathrm{CT}}(\mathbf{w}) = L_{\mathrm{CT}}(\mathbf{w}) - \beta\phi(\mathbf{w}) - \gamma\rho(\mathbf{w}) \quad (3)$$

$$\Phi_{\mathrm{BM}}(\mathbf{w}) = L_{\mathrm{BM}}(\mathbf{w}) - \beta\phi(\mathbf{w}) - \gamma\rho(\mathbf{w}) \quad (4)$$

where β is the spatial hyperparameter, γ is the temporal hyperparameter, and $\phi(\mathbf{w})$ and $\rho(\mathbf{w})$ are the spatial and temporal penalties, respectively, given by [16]

$$\phi(\mathbf{w}) = \sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v}\sum_{j'\in\mathcal{N}_j, j'>j} \zeta_{jj'}(w_{j\ell} - w_{j'\ell})^2 \quad (5)$$

$$\rho(\mathbf{w}) = \sum_{j=1}^{n_v}\sum_{\ell_1=1}^{n_b}\sum_{\ell_2=1}^{n_b} w_{j\ell_1}Z_{\ell_1\ell_2}w_{j\ell_2} \quad (6)$$

where $\zeta_{jj'}$ is the reciprocal Euclidean distance between voxels j and $j', \mathcal{N}_j$ denotes the neighborhood of voxel j, and $\mathbf{Z}$ is the symmetric banded matrix of the quadratic form used to express cubic B-splines' integrated squared curvature [21].

Under the inhomogeneous Poisson process model, the (i,n)th element of our observation vector $\mathbf{y}$, (i.e., $y_i^{(n)}$) has a Poisson distribution with mean $\int_{t_{n-1}}^{t_n} r_i(t)\,dt$ where t_{n-1} and t_n denote the end points of time bin n and $r_i(t)$ denotes the true rate function at detector pair i. By substituting $\sum_\ell\sum_j p_{ij}w_{j\ell}B_\ell(t)$ for $r_i(t)$, we can relate the mean number of events at each spatiotemporal bin $\bar{y}_i^{(n)}$ to the parameters to be estimated as follows:

$$\bar{y}_i^{(n)} = \sum_{j=1}^{n_v}\sum_{\ell=1}^{n_b} p_{ij}\left[\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right] w_{j\ell}. \quad (7)$$

Note the space-time separability in (7) which allows us to write the mean of our observation vector in terms of the control vertex vector $\mathbf{w}$ as follows:

$$\bar{\mathbf{y}} = (\mathbf{P}\otimes\mathbf{B})\mathbf{w}. \quad (8)$$

Here, $\mathbf{B}$ is the temporal sensitivity matrix whose (n,l)th element is given by $\int_{t_{n-1}}^{t_n} B_l(t)\,dt$ and $\otimes$ denotes the left Kronecker product. This space-time separability in the imaging equation is the combined result of a temporally constant system matrix and the use of the same temporal basis functions at each voxel/detector pair. We note that this separability assumes that the system matrix is time invariant and hence does not allow for spatially and temporally variant deadtime models, although a system wide spatially variant deadtime factor can be accounted for after reconstruction. Also note that this bin-mode imaging equation has the same form as the static imaging equation $\bar{\mathbf{y}} = \mathbf{P}\mathbf{x}$ except that the system matrix $\mathbf{P}$, image vector $\mathbf{x}$, and mean sinogram data $\bar{\mathbf{y}}$ are replaced, respectively, by the spatiotemporal system matrix $\mathbf{P}\otimes\mathbf{B}$, control vertex vector $\mathbf{w}$, and mean bin-mode data also denoted by $\bar{\mathbf{y}}$. We make use of this mathematical equivalence in the following sections in deriving our mean and variance approximations.

B. Mean Estimation

A simple but reasonably accurate method for estimating the mean reconstructed image in static (i.e., single frame) reconstructions is to reconstruct noise-free data [5]. Even if the means for the total number of detected events at each detector pair are nonintegers, they can be formally substituted into the static PET objective function and the optimization of that objective function will give an estimate for the mean reconstructed image.

The same approach in the dynamic case is not possible with the continuous time log-likelihood function because it is based on the event arrival times rather than the total number of events at each detector pair. However, the bin-mode approach allows for a similar substitution in which the $y_i^{(n)}$'s are replaced by (possibly noninteger) $\bar{y}_i^{(n)} = \int_{t_{n-1}}^{t_n} r_i(t)\,dt$. Substitution of this into (2) results in

$$L_{\mathrm{BM}}(\mathbf{w}) = \sum_{i=1}^{n_d}\sum_{n=1}^{N}\bar{y}_i^{(n)} \log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right) - \sum_{i=1}^{n_d}\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_0^T B_\ell(t)\,dt\right). \quad (9)$$

If the temporal basis functions are differentiable, the mean value theorem allows us to write $\bar{y}_i^{(n)} = r_i(c_{\mathrm{in}})(t_n - t_{n-1})$ and $\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt = B_\ell(c'_{\ell n})(t_n - t_{n-1})$ where $t_{n-1} \le c_{\mathrm{in}} \le t_n$ and $t_{n-1} \le c'_{\ell n} \le t_n$. These substitutions result in

$$\begin{aligned} L_{\mathrm{BM}}(\mathbf{w}) &= \sum_{i=1}^{n_d}\sum_{n=1}^{N} r_i(c_{\mathrm{in}})\Delta n \log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}B_\ell(c'_{\ell n})\Delta n\right) \\ &\quad - \sum_{i=1}^{n_d}\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_0^T B_\ell(t)\,dt \end{aligned} \quad (10)$$

where $\Delta n = t_n - t_{n-1}$. Removing the terms that do not depend on $\mathbf{w}$ and allowing all bin widths to approach zero, we obtain

$$L(\mathbf{r}(t), \mathbf{w}) = \sum_{i=1}^{n_d}\int_0^T r_i(t)\log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}B_\ell(t)\right)dt - \sum_{i=1}^{n_d}\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_0^T B_\ell(t)\,dt \quad (11)$$

where $\mathbf{r}(t)$ is a vector valued function of time whose ith element is $r_i(t)$. Together with the penalty terms, our implicitly defined mean approximation then becomes

$$E\{\hat{\mathbf{w}}\} \approx \arg\max_{\mathbf{w}\in\mathcal{W}} L(\mathbf{r}(t), \mathbf{w}) - \beta\phi(\mathbf{w}) - \gamma\rho(\mathbf{w}) \quad (12)$$

where $\mathcal{W} = \{\mathbf{w} \mid \sum_{\ell=1}^{n_b} w_{j\ell}B_\ell(t) \ge 0\ \forall j, t$ and $\sum_{j=1}^{n_v}\sum_{\ell=1}^{n_b} p_{ij}w_{j\ell}B_\ell(a_{ik}) > 0\ \forall\ i, k\}$. Although (12) does not provide a closed form approximation to the mean reconstructed image, it requires only a single reconstruction using the rate functions at detector pairs, compared to the hundreds of noisy reconstructions that are necessary for Monte Carlo simulations. We note that reconstruction of the mean image takes approximately 5 times longer than reconstructions from noisy data for realistic count rates, since in the latter case a large fraction of the sinogram bins are empty (zero counts), while the mean of the data is nonzero. However, the computational cost is still considerably lower than that of using Monte Carlo techniques.

We also note that [compare (12) with (1)] the reconstruction of a dynamic image from event arrival times and our mean approximation both fall into the general category of maximizing an objective function of the form

$$\Phi(\mathbf{f}(t), \mathbf{w}) = \sum_{i=1}^{n_d}\int_0^T f_i(t)\log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}B_\ell(t)\right)dt - \sum_{i=1}^{n_d}\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij}w_{j\ell}\int_0^T B_\ell(t)\,dt - \beta\phi(\mathbf{w}) - \gamma\rho(\mathbf{w}) \quad (13)$$

where $\mathbf{f}(t)$ is a vector valued function of time whose ith element is given by $f_i(t) = r_i(t)$ in the "noiseless" case and $f_i(t) = \sum_{k=1}^{x_i}\delta(t - a_{ik})$ in the "noisy" case where $\delta(\cdot)$ denotes the Dirac delta function. This result is intuitive in the sense that when the rate functions at detector pairs are known (noiseless), a reconstruction using those rate functions gives a first order estimate for the mean. On the other hand, we can think of the noisy situation (i.e., when we have a realization of the inhomogeneous Poisson process instead of its rate function) as one in which the rate functions $r_i(t)$ are replaced by their estimates $\sum_{k=1}^{x_i}\delta(t - a_{ik})$ corresponding to an impulse train at event arrival times.

C. Variance Estimation

Instantaneous rate estimates represent the activity at a given voxel at a fixed point in time, whereas average rate estimates represent the activity averaged over a finite time interval. Under our parameterized inhomogeneous Poisson process model, the covariance between rate function estimates at any two voxels j and k at any two time points t_m and t_n (instantaneous) or averaged from t_1 to t_2 (average) can be obtained via

$$\begin{aligned} \mathrm{cov}\,(\hat{\eta}_j(t_m), \hat{\eta}_k(t_n)) &= \mathrm{cov}\left(\sum_{\ell=1}^{n_b}\hat{w}_{j\ell}B_\ell(t_m), \sum_{\ell=1}^{n_b}\hat{w}_{k\ell}B_\ell(t_n)\right) \\ &= \mathbf{b}(t_m)^T\mathbf{C}_{\hat{w}_j,\hat{w}_k}\mathbf{b}(t_n) \end{aligned} \quad (14)$$

$$\mathrm{cov}\left(\frac{\int_{t_1}^{t_2}\hat{\eta}_j(t)\,dt}{(t_2 - t_1)}, \frac{\int_{t_1}^{t_2}\hat{\eta}_k(t)\,dt}{(t_2 - t_1)}\right) = \frac{1}{(t_2 - t_1)^2}\left(\int_{t_1}^{t_2}\mathbf{b}(t)\,dt\right)^T\mathbf{C}_{\hat{w}_j,\hat{w}_k}\left(\int_{t_1}^{t_2}\mathbf{b}(t)\,dt\right) \quad (15)$$

where

$$\mathbf{b}(t) \equiv [B_1(t)\ldots B_{n_b}(t)]^T \quad (16)$$

$$\mathbf{C}_{\hat{w}_j,\hat{w}_k} \equiv E[(\hat{\mathbf{w}}_j - E(\hat{\mathbf{w}}_j))(\hat{\mathbf{w}}_k - E(\hat{\mathbf{w}}_k))^T]. \quad (17)$$

Since rate function estimates are related to control vertices deterministically, an estimate of the covariance matrix, $\mathbf{C}_{\mathbf{w}} \equiv E[(\hat{\mathbf{w}} - E(\hat{\mathbf{w}}))(\hat{\mathbf{w}} - E(\hat{\mathbf{w}}))^T]$ is necessary and sufficient for estimating the variances/covariances of instantaneous or average rate estimates at any pair of voxels at any pair of time points. Once $\mathbf{C}_{\mathbf{w}}$ is estimated, it can be used in (14) or (15) to obtain the desired expressions.

D. Covariance Matrix Estimation

In this subsection we will derive an approximation to $\mathbf{C}_{\mathbf{w}}$ in the context of penalized ML estimation and show how it can be evaluated efficiently using discrete Fourier transforms (DFTs). We start by using the bin-mode imaging equation and the fact that both spatial and temporal penalties are quadratic to derive the approximate covariance expression. We then make further approximations so that the resulting covariance expression only requires the inversion of circulant or block circulant matrices that can be performed efficiently using DFTs.

We showed in Section II-A that the bin-mode imaging equation could be expressed as $\bar{\mathbf{y}} = (\mathbf{P} \otimes \mathbf{B})\mathbf{w}$. The spatial and temporal smoothness penalties used in the reconstruction are both quadratic in the control vertices and, therefore, can be put in the standard quadratic form

$$\phi(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{R}_S\mathbf{w} \tag{18}$$

$$\rho(\mathbf{w}) = \frac{1}{2}\mathbf{w}^T\mathbf{R}_T\mathbf{w}. \tag{19}$$

The bin-mode imaging equation (8) together with the quadratic spatial and temporal penalties allows us to adopt approximations similar to those in [5] to achieve the following approximation for $\mathbf{C_w}$:

$$\mathbf{C_w} \approx (\mathbf{F} + \beta\mathbf{R}_S + \gamma\mathbf{R}_T)^{-1}\mathbf{F}(\mathbf{F} + \beta\mathbf{R}_S + \gamma\mathbf{R}_T)^{-1} \tag{20}$$

where $\mathbf{F}$ is the FIM given by

$$\mathbf{F} \equiv (\mathbf{P} \otimes \mathbf{B})^T\mathbf{D}\left\{\frac{1}{\bar{y}_i^{(n)}}\right\}(\mathbf{P} \otimes \mathbf{B}) \tag{21}$$

where $\mathbf{D}\{\cdot\}$ denotes a diagonal matrix. The difficulty in evaluating the covariance matrix expression (20) is in the matrix inversions and we will use Fourier transform arguments as in [22] and [8] together with properties of Kronecker products [9] to obtain an approximation to $\mathbf{C_w}$ that can be computed efficiently.

First we introduce the double index notation $j\ell$ to denote $((j-1)n_b + \ell)$ and note that the $(j\ell, j'\ell')$th element of the $n_b n_v \times n_b n_v$ FIM is given by

$$\mathbf{F}_{j\ell,j'\ell'} = \sum_{i=1}^{n_d}\sum_{n=1}^{N}\frac{p_{ij}p_{ij'}B_{n\ell}B_{n'\ell'}}{\bar{y}_i^{(n)}}. \tag{22}$$

Using this exact form of the FIM together with the penalty terms in (18) and (19) makes the computation of $(\mathbf{F} + \beta\mathbf{R}_S + \gamma\mathbf{R}_T)^{-1}$ very difficult. Therefore, we proceed by making an approximation to the FIM very similar to that in [5]

$$\mathbf{F} \approx \mathbf{D}_\kappa\mathbf{D}_v^{-1}(\mathbf{P} \otimes \mathbf{B})^T(\mathbf{P} \otimes \mathbf{B})\mathbf{D}_v^{-1}\mathbf{D}_\kappa \tag{23}$$

where $\mathbf{D}_\kappa \equiv \mathbf{D}\{\kappa_{j\ell}\}, \mathbf{D}_v \equiv \mathbf{D}\{v_{j\ell}\}$. Here, $\kappa_{j\ell}$ and $v_{j\ell}$ are given by

$$\kappa_{j\ell} = \sqrt{\sum_{i=1}^{n_d} p_{ij}^2\left(\int_0^T \frac{B_\ell^2(t)\,dt}{r_i(t)}\right)} \tag{24}$$

$$v_{j\ell} = \sqrt{\sum_{i=1}^{n_d} p_{ij}^2\sum_{n=1}^{N}\left(\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right)^2} \tag{25}$$

where $r_i(t)$ denotes the rate function at the ith detector pair. Note that the $\kappa_{j\ell}$'s are the square roots of the diagonal elements of the dynamic FIM (i.e., $\kappa_{j\ell}^2 = \mathbf{F}_{j\ell,j\ell}$) as the bin-width approaches zero

$$\begin{aligned}\mathbf{F}_{j\ell,j\ell} &= \sum_{i=1}^{n_d}\sum_{n=1}^{N}p_{ij}^2\frac{\left(\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right)^2}{\bar{y}_i^{(n)}} \\ &\approx \sum_{i=1}^{n_d}\sum_{n=1}^{N}p_{ij}^2\frac{B_\ell^2(t_n)(\Delta n)^2}{r_i(t_n)(\Delta n)} \\ &= \sum_{i=1}^{n_d}p_{ij}^2\sum_{n=1}^{N}\frac{B_\ell^2(t_n)(\Delta n)}{r_i(t_n)}\end{aligned} \tag{26}$$

$$\lim_{N\to\infty}\mathbf{F}_{j\ell,j\ell} = \sum_{i=1}^{n_d}p_{ij}^2\int_0^T\frac{B_\ell^2(t)\,dt}{r_i(t)}. \tag{27}$$

Similarly, the $v_{j\ell}$'s are square roots of "spatiotemporal backprojections" of uniform data (i.e., $y_i^{(n)} = 1\ \forall i, n$) where the backprojection is performed with a matrix which is the elementwise square of $(\mathbf{P} \otimes \mathbf{B})$

$$\begin{aligned}v_{j\ell}^2 &= \sum_{i=1}^{n_d}\sum_{n=1}^{N}[(\mathbf{P} \otimes \mathbf{B})_{\text{in},j\ell}]^2 \cdot 1 \\ &= \sum_{i=1}^{n_d}\sum_{n=1}^{N}\left[p_{ij}\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right]^2 \\ &= \sum_{i=1}^{n_d}p_{ij}^2\sum_{n=1}^{N}\left(\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right)^2.\end{aligned} \tag{28}$$

We also note that based on (26), (24) needs to be computed only over detector pairs that have nonzero probabilities of detecting activity at voxel j and only over the supports of the basis functions. Hence, we have

$$\kappa_{j\ell} = \sqrt{\sum_{i\in\mathcal{I}_j}p_{ij}^2\left(\int_{t\in\mathcal{T}_\ell}\frac{B_\ell^2(t)\,dt}{r_i(t)}\right)} \tag{29}$$

where $\mathcal{I}_j = \{i\,|\,p_{ij} > 0\}$ and $\mathcal{T}_\ell = \{t | B_\ell(t) > 0\}$. Therefore, if $r_{i'}(t) = 0$ over an interval $\Omega \subset [0, T]$ and $p_{i'j} > 0$ and $B_\ell(t) > 0$ over a subset of Ω then $\kappa_{j\ell} = \infty$. Otherwise the values of $r_{i'}(t)$ over Ω do not contribute to the FIM computation.

We proceed by noting the Kronecker identity $(\mathbf{P} \otimes \mathbf{B})^T(\mathbf{P} \otimes \mathbf{B}) = (\mathbf{P^T P} \otimes \mathbf{B}^T\mathbf{B})$ and observing that $v_{j\ell}$ is separable into $v_{j\ell} = v_j^{(s)}v_\ell^{(t)}$ where

$$v_j^{(s)} = \sqrt{\sum_{i=1}^{n_d}p_{ij}^2} \quad v_\ell^{(t)} = \sqrt{\sum_{n=1}^{N}\left(\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right)^2}. \tag{30}$$

Superscripts "(s)" and "(t)" are used to emphasize the fact that these are spatial and temporal terms, respectively. This separability leads to the factorization $\mathbf{D}\{v_{j\ell}\} = (\mathbf{D}\{v_j^{(s)}\} \otimes \mathbf{D}\{v_\ell^{(t)}\})$ and our FIM approximation becomes

$$\begin{aligned}\mathbf{F} &\approx \mathbf{D}_\kappa\left(\mathbf{D}_s \otimes \mathbf{D}_t\right)^{-1}\left(\mathbf{P}^T\mathbf{P} \otimes \mathbf{B}^T\mathbf{B}\right)\left(\mathbf{D}_s \otimes \mathbf{D}_t\right)^{-1}\mathbf{D}_\kappa \\ &= \mathbf{D}_\kappa\left[\left[\left(\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}\right)\right] \otimes \left[\mathbf{D}_t{}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{D}_t{}^{-1}\right)\right]\right]\mathbf{D}_\kappa\end{aligned} \tag{31}$$

where $\mathbf{D}_s \equiv \mathbf{D}\{v_j^{(s)}\}, \mathbf{D}_t \equiv \mathbf{D}\{v_\ell^{(t)}\}$ and we used the properties of the Kronecker product to write the second part of (31).

Now we turn our attention to the other two terms in the covariance matrix (20), $\beta\mathbf{R}_S$ and $\gamma\mathbf{R}_T$. It follows from (5)–(19) that we can rewrite $\mathbf{R}_S$ and $\mathbf{R}_T$ as

$$\mathbf{R}_S = \mathbf{S}\otimes\mathbf{I}_b \tag{32}$$
$$\mathbf{R}_T = \mathbf{I}_v\otimes\mathbf{T} \tag{33}$$

where $T_{\ell'\ell''} \equiv 2Z_{\ell'\ell''}$ and $S_{j'j''} \equiv 2\sum_{j\in N_j}\zeta_{jj'}$ if $j'=j''$ and $-2\zeta_{j'j''}$ if $j'\neq j''$. The identity matrix $\mathbf{I}_b$ in (32) is $n_b\times n_b$ and the identity matrix $\mathbf{I}_v$ in (33) is $n_v\times n_v$. These equations show that both the spatial and temporal penalties are separable in space and time. This separability is a result of the application of the same spatial penalty over time (at every control vertex ℓ, the spatial penalty is the squared difference between neighboring voxels in the control vertex image) and the application of the same temporal penalty over space (at every voxel j, the temporal penalty is the integrated square curvature). We proceed by expressing $\beta\mathbf{R}_S$ and $\gamma\mathbf{R}_T$ as follows:

$$\beta\mathbf{R}_S = \mathbf{D}_\kappa\mathbf{D}_\eta(\mathbf{S}\otimes\mathbf{I})\mathbf{D}_\eta\mathbf{D}_\kappa \tag{34}$$
$$\gamma\mathbf{R}_T = \mathbf{D}_\kappa\mathbf{D}_\psi(\mathbf{I}\otimes\mathbf{T})\mathbf{D}_\psi\mathbf{D}_\kappa \tag{35}$$

where $\mathbf{D}_\eta\equiv\mathbf{D}\{\eta_{j\ell}\},\mathbf{D}_\psi\equiv\mathbf{D}\{\psi_{j\ell}\}$ and $\eta_{j\ell}$ and $\psi_{j\ell}$ are given by

$$\eta_{j\ell} = \sqrt{\beta}/\kappa_{j\ell} \tag{36}$$
$$\psi_{j\ell} = \sqrt{\gamma}/\kappa_{j\ell}. \tag{37}$$

We can now express $\mathbf{C_w}$ in the following form:

$$\mathbf{C_w}\approx[\mathbf{F}+\mathbf{D}_\kappa\mathbf{D}_\eta(\mathbf{S}\otimes\mathbf{I})\mathbf{D}_\eta\mathbf{D}_\kappa+\mathbf{D}_\kappa\mathbf{D}_\psi(\mathbf{I}\otimes\mathbf{T})\mathbf{D}_\psi\mathbf{D}_\kappa]^{-1}$$
$$\mathbf{F}[\mathbf{F}+\mathbf{D}_\kappa\mathbf{D}_\eta(\mathbf{S}\otimes\mathbf{I})\mathbf{D}_\eta\mathbf{D}_\kappa+\mathbf{D}_\kappa\mathbf{D}_\psi(\mathbf{I}\otimes\mathbf{T})\mathbf{D}_\psi\mathbf{D}_\kappa]^{-1} \tag{38}$$

where $\mathbf{F}$ is approximated by (31). In order to be able to diagonalize the remaining terms after factoring out $\mathbf{D}_\kappa$ we make the following approximation, similar to its static counterpart in [8]

$$\begin{aligned}&\left[\left(\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}\right)\otimes\left(\mathbf{D}_t{}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{D}_t{}^{-1}\right)\right.\\&\quad\left.+\,\mathbf{D}_\eta(\mathbf{S}\otimes\mathbf{I})\mathbf{D}_\eta+\mathbf{D}_\psi(\mathbf{I}\otimes\mathbf{T})\mathbf{D}_\psi\right]\mathbf{e}_{j\ell}\\&\qquad\approx\left[\left(\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}\right)\otimes\left(\mathbf{D}_t{}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{D}_t{}^{-1}\right)\right.\\&\qquad\quad\left.+\eta_{j\ell}^2(\mathbf{S}\otimes\mathbf{I})+\psi_{j\ell}^2(\mathbf{I}\otimes\mathbf{T})\right]\mathbf{e}_{j\ell}\end{aligned} \tag{39}$$

where $\mathbf{e}_{j\ell}$ denotes the unit vector in the $j\ell$th direction. The accuracy of this approximation depends on how slowly the $\eta_{j\ell}$'s and $\psi_{j\ell}$'s vary and the approximation becomes exact when all $\eta_{j\ell}$'s and $\psi_{j\ell}$'s are equal. Note that the $\kappa_{j\ell}$'s are smooth functions of mean activity at detector pairs due to the integration and backprojection operations in (24). Therefore, the $\eta_{j\ell}$'s and $\psi_{j\ell}$'s are also expected to change slowly over the dynamic image [see (36) and (37)] unless the $\kappa_{j\ell}$'s are very small. As a result, our covariance approximation is given by (40), shown at the bottom of the page, where

$$\begin{aligned}\mathbf{A}_1\equiv&\left[\left(\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}\right)\otimes\left(\mathbf{D}_t{}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{D}_t{}^{-1}\right)\right.\\&\left.+\,\eta_{j'\ell'}^2(\mathbf{S}\otimes\mathbf{I})+\psi_{j'\ell'}^2(\mathbf{I}\otimes\mathbf{T})\right]^{-1}\end{aligned} \tag{41}$$
$$\begin{aligned}\mathbf{A}_2\equiv&\left[\left(\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}\right)\otimes\left(\mathbf{D}_t{}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{D}_t{}^{-1}\right)\right.\\&\left.+\eta_{j''\ell''}^2(\mathbf{S}\otimes\mathbf{I})+\psi_{j''\ell''}^2(\mathbf{I}\otimes\mathbf{T})\right]^{-1}.\end{aligned} \tag{42}$$

Inspection of (40) is our final approximation for the covariance between two control vertices. In the next subsection we describe how it can be evaluated efficiently using DFTs.

E. Efficient Evaluation of the Covariance Matrix

An inspection of (40)–(42) shows that if we could diagonalize $\mathbf{P}^T\mathbf{P},\mathbf{B}^T\mathbf{B},\mathbf{S}$, and $\mathbf{T}$, then we could evaluate the covariance efficiently because the other matrices involved in the expression are already diagonal. Therefore, we approximately diagonalize $\mathbf{P}^T\mathbf{P}$ and $\mathbf{S}$ using two-dimensional (2-D) or three-dimensional (3-D) 3D-DFTs (2D-DFTs for 2-D reconstructions and 3D-DFTs for fully 3-D reconstructions) based on the assumption that they are approximately doubly block circulant. The accuracy of the diagonalizations has been demonstrated in the analysis of the static PET problem [8]. Here, $\mathbf{P}$ includes only the geometric detection probabilities but components of $\mathbf{P}$ corresponding to attenuation and normalization can also be incorporated to the approximation as in [8]. We also diagonalize $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ using one–dimensional DFTs (1D-DFTs) assuming that they are approximately circulant. Any $n_b\times n_b$ unitary matrix which can approximately diagonalize both $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ can also be used in the diagonalization. Approximate diagonalization leads to the following expressions:

$$\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}\approx\mathbf{Q}_v^T\mathbf{D}\left\{\lambda_j^{(s)}\right\}\mathbf{B}^T\mathbf{Q}_v \tag{43}$$
$$\mathbf{D}_t{}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{D}_t{}^{-1}\approx\mathbf{Q}_b{}^T\mathbf{D}\left\{\lambda_\ell^{(t)}\right\}\mathbf{Q}_b \tag{44}$$
$$\mathbf{S}\approx\mathbf{Q}_v^T\mathbf{B}^T\mathbf{D}\{\mu_j\}\mathbf{Q}_v \tag{45}$$
$$\mathbf{T}\approx\mathbf{Q}_b{}^T\mathbf{D}\{\delta_\ell\}\mathbf{Q}_b \tag{46}$$

where $\mathbf{Q}_v$ denotes the Kronecker form of the $n_v\times n_v$ 2D-DFT matrix and $\mathbf{Q}_b$ denotes the $n_b\times n_b$ 1D-DFT matrix [23].

The mathematical background behind these diagonalizations is that circulant matrices can be diagonalized exactly and that the diagonalizing matrix is the DFT matrix of appropriate dimensions. Therefore, $\{\lambda_j^{(s)}\}_{j=1}^{n_v}$ and $\{\mu_j\}_{j=1}^{n_v}$ are computed by taking the 3D-DFTs of the 3-D matrices (or 2D-DFTs of the 2-D matrices) formed by reshaping the central columns of $\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}$ and $\mathbf{S}$ into 3-D (or 2D) matrices. Similarly, $\{\lambda_\ell^{(t)}\}_{\ell=1}^{n_b}$ and $\{\delta_\ell\}_{\ell=1}^{n_b}$ are computed by taking the 1D-DFTs

$$\begin{aligned}\mathrm{cov}(\hat{w}_{j'\ell'},\hat{w}_{j''\ell''})=&\,\mathbf{e}_{j'\ell'}^T\mathbf{C_w}\mathbf{e}_{j''\ell''}\\\approx&\,\frac{\mathbf{e}_{j'\ell'}^T\mathbf{A}_1\left[\left(\mathbf{D}_s{}^{-1}\mathbf{P}^T\mathbf{P}\mathbf{D}_s{}^{-1}\right)\otimes\mathbf{D}_t{}^{-1}\mathbf{B}^T\mathbf{B}\mathbf{D}_t{}^{-1}\right)\right]\mathbf{A}_2\mathbf{e}_{j''\ell''}}{\kappa_{j'\ell'}\kappa_{j''\ell''}}\end{aligned} \tag{40}$$

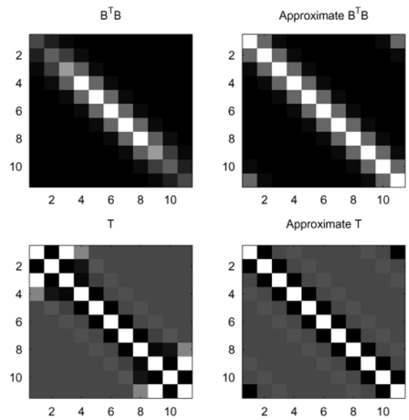

Fig. 1. Original $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ matrices and their circulant approximations for a uniform B-spline basis.

of ${\mathbf{D}_t}^{-1}\mathbf{B}^T\mathbf{B}{\mathbf{D}_t}^{-1}$ and $\mathbf{T}$. These circulant matrix approximations are equivalent to assuming that the columns of the diagonalized matrices are approximately shifted replicas of each other. Figs. 1 and 2 demonstrate circulant approximations for $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ for uniform and nonuniform B-splines, respectively. In cases where circulant assumptions do not hold, the approximation accuracy can be improved by taking the DFTs of the columns that correspond to the voxel (or control vertex) of interest [8].

We also note that the $\{\lambda_j^{(s)}\}_{j=1}^{n_v}$ and $\{\mu_j\}_{j=1}^{n_v}$ are exactly the same 3D-DFT (or 2D-DFT) coefficients used in the static PET approximations [8]. Although $\mathbf{D}_t$ and $\mathbf{B}^T\mathbf{B}$ are functions of the bin-width, as the bin-width approaches zero, the ℓkth element of ${\mathbf{D}_t}^{-1}\mathbf{B}^T\mathbf{B}{\mathbf{D}_t}^{-1}$ is given by

$$\begin{aligned}\left[{\mathbf{D}_t}^{-1}\mathbf{B}^T\mathbf{B}{\mathbf{D}_t}^{-1}\right]_{\ell,k} &= \frac{\sum_{n=1}^{N}\left(\int_{t_{n-1}}^{t_n} B_\ell(t)\,dt\right)\left(\int_{t_{n-1}}^{t_n} B_k(t)\,dt\right)}{\sum_{n=1}^{N}\left(\int_{t_{n-1}}^{t_n} B_k(t)\,dt\right)^2}\\ &\approx \frac{\sum_{n=1}^{N} B_\ell(t_n)B_k(t_n)(\Delta n)}{\sum_{n=1}^{N} B_k^2(t_n)(\Delta n)} \qquad (47)\end{aligned}$$

$$\begin{aligned}\lim_{N\to\infty}\left[{\mathbf{D}_t}^{-1}\mathbf{B}^T\mathbf{B}{\mathbf{D}_t}^{-1}\right]_{\ell,k} &= \frac{\int_0^T B_\ell(t)B_k(t)\,dt}{\int_0^T B_\ell^2(t)\,dt} \qquad (48)\end{aligned}$$

and, therefore, *none* of the diagonalizations require the list-mode data to be sorted into small spatiotemporal bins. Substitution of these diagonalizations into (40) results in the following covariance approximation:

$$\operatorname{cov}(\hat{w}_{j'\ell'},\hat{w}_{j''\ell''}) = \frac{\mathbf{e}_{j'\ell'}^T\mathbf{Q}^T\mathbf{D}\mathbf{Q}\mathbf{e}_{j''\ell''}}{\kappa_{j'\ell'}\kappa_{j''\ell''}} \qquad (49)$$

where $\mathbf{Q}\equiv\mathbf{Q}_v\otimes\mathbf{Q}_b$ is the 4D-DFT matrix (3D-DFT matrix for two spatial dimensions) and $\mathbf{D}$ is a diagonal matrix whose diagonal entries are given by

$$\mathbf{D}_{j\ell,j\ell} = \frac{\lambda_j^{(s)}\lambda_\ell^{(t)}}{\left[\lambda_j^{(s)}\lambda_\ell^{(t)}+\kappa_{j'\ell'}^{-2}(\beta\mu_j+\gamma\delta_\ell)\right]\left[\lambda_j^{(s)}\lambda_\ell^{(t)}+\kappa_{j''\ell''}^{-2}(\beta\mu_j+\gamma\delta_\ell)\right]}. \qquad (50)$$

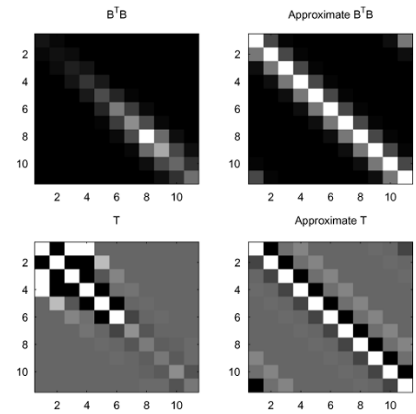

Fig. 2. Original $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ matrices and their circulant approximations for a nonuniform B-spline basis. Knot spacings were chosen at equal arc lengths along the "total activity versus time" curve for the simulation in Section III.

The calculation of the full $\mathbf{Q}$ matrix is not necessary in computing the covariance approximation in (49). The (j,ℓ)th column of $\mathbf{Q}$ can be computed by taking the Kronecker product of the jth column of $\mathbf{Q}_v$ with the ℓth column of $\mathbf{Q}_b$. When variances are calculated (i.e., $j'=j''$ and $\ell'=\ell''$) the unitary property of $\mathbf{Q}$ allows for the following simplified approximate variance expression:

$$\operatorname{var}(\hat{w}_{j'\ell'}) = \frac{\kappa_{j'\ell'}^{-2}}{n_b n_v}\sum_{j=1}^{n_v}\sum_{\ell=1}^{n_b}\frac{\lambda_j^{(s)}\lambda_\ell^{(t)}}{\left(\lambda_j^{(s)}\lambda_\ell^{(t)}+\kappa_{j'\ell'}^{-2}(\beta\mu_j+\gamma\delta_\ell)\right)^2}. \qquad (51)$$

Note that many of the spatial approximations used in the derivation of (49)–(51) derive from the literature [5]–[8] and the circulant approximations of temporal matrices are demonstrated in Figs. 1 and 2. In Section III we present computational examples to illustrate the overall accuracy of our covariance approximation. In these examples, our covariance approximation provides accurate estimates despite spatiotemporal activity differences up to an order of magnitude.

It is important to note that list-mode data does *not* need to be sorted into very small spatiotemporal bins for these approximations to hold. By viewing penalized ML continuous-time PET images reconstructed from continuous-time list-mode data as the limiting case of those reconstructed from bin-mode data, we can view (12) and (49) as efficient approximations of the mean and covariance for continuous-time reconstructions.

F. Comparison With Static Approximations

In this subsection we simplify (51) for the static case and compare the resulting expression with the result of the derivation in [8]. In the static case, $n_b=1, N=1$ and $B_1(t)=(1)/(T)\forall t\in[0,T]$ because we have a single, constant basis function that integrates to 1 and consists of a single frame. Furthermore, since there is no temporal penalty, $\gamma\equiv 0$. These simplifications lead to the following:

$$\mathbf{B}^T\mathbf{B} = \left[\int_0^T B_1(t)\,dt\right] = \left[\int_0^T \frac{1}{T}\,dt\right] = [1] \Rightarrow \lambda_1^{(t)} = 1 \qquad (52)$$

$$v_{j\ell} \equiv v_j = \sqrt{\sum_{i=1}^{n_d} p_{ij}^2 \left(\int_0^T \frac{1}{T}\, dt \right)^2} = \sqrt{\sum_{i=1}^{n_d} p_{ij}^2} \tag{53}$$

$$\begin{aligned} \kappa_{j\ell} \equiv \kappa_j &= \sqrt{\sum_{n=1}^{N} \sum_{i=1}^{n_d} p_{ij}^2 \frac{\left(\int_{t_{n-1}}^{t_n} B_1(t)\, dt \right)^2}{\bar{y}_i^{(n)}}} \\ &= \sqrt{\sum_{i=1}^{n_d} p_{ij}^2 \frac{\left(\int_0^T \frac{1}{T}\, dt \right)^2}{\bar{y}_i}} \\ &= \sqrt{\sum_{i=1}^{n_d} \frac{p_{ij}^2}{\bar{y}_i}}. \end{aligned} \tag{54}$$

Substituting (52)–(54) into (49) leads to

$$\mathrm{var}(\hat{w}_{j'}) = \frac{\kappa_{j'}^{-2}}{n_v} \sum_{j=1}^{n_v} \frac{\lambda_j^{(s)}}{\left(\lambda_j^{(s)} + \kappa_{j'}^{-2} \beta \mu_j \right)^2} \tag{55}$$

where κ_j is given by (54) and the $\lambda_j^{(s)}$ and μ_j are given by (43) and (45). This is exactly the same expression as the variance formula in [8] and shows that the approximations given in this section can be viewed as a consistent extension of the static PET mean and variance approximations to dynamic PET reconstructions.

III. Simulations

In order to evaluate our mean and covariance approximations we simulated a single-ring scanner with the parameters of the ECAT HR+ clinical scanner. Image and sinogram dimensions were 128×128 and 288×144, respectively. Eleven cubic B-splines with uniformly spaced knots were used as temporal basis functions resulting in $128 \times 128 \times 11$ spatiotemporal coefficients. The sinogram portion of sorted list-mode data was simulated by generating a Poisson random vector $\mathbf{y}$, with mean equal to the forward projection of the simulated image $\mathbf{x}_{\mathrm{SIM}}$ (i.e., $\bar{\mathbf{y}} = \mathbf{P}\mathbf{x}_{\mathrm{SIM}}$). The "timogram" [16] portion of list-mode data was simulated by generating y_i uniform random variables between $[0, T]$ for each sinogram bin i and warping the resulting arrival times according to the inverse of the integrated rate function at that bin.

We ran 30 iterations of our dynamic conjugate gradient MAP algorithm [16] that achieved effective convergence in all reconstructions with uniform B-splines. In reconstructions with nonuniform B-splines, we used 100 conjugate gradient iterations. Attenuation and normalization effects were not included and the positivity constraints were inactive at the voxels of interest. Three different rate functions were simulated to represent activity in the gray matter, white matter and cerebrospinal fluid (CSF). Fig. 3 shows the true sixth control vertex image. Temporal basis functions and simulated rate functions are plotted in Fig. 4.

Exact κ's [i.e., exact $r_i(t)'s$ used in (24)] were used in the computation of (49). Note that both our mean and covariance approximations require that the mean of the data be known. The evaluation of (12) and (49) by estimating mean data from observed data will require an extension of the static FIM estimation techniques described in [20].

Fig. 3. The true image for the sixth control vertex illustrating three different regions for white matter, gray matter and cerebrospinal fluid on the Hoffman phantom. The black line corresponds to the central horizontal profile.

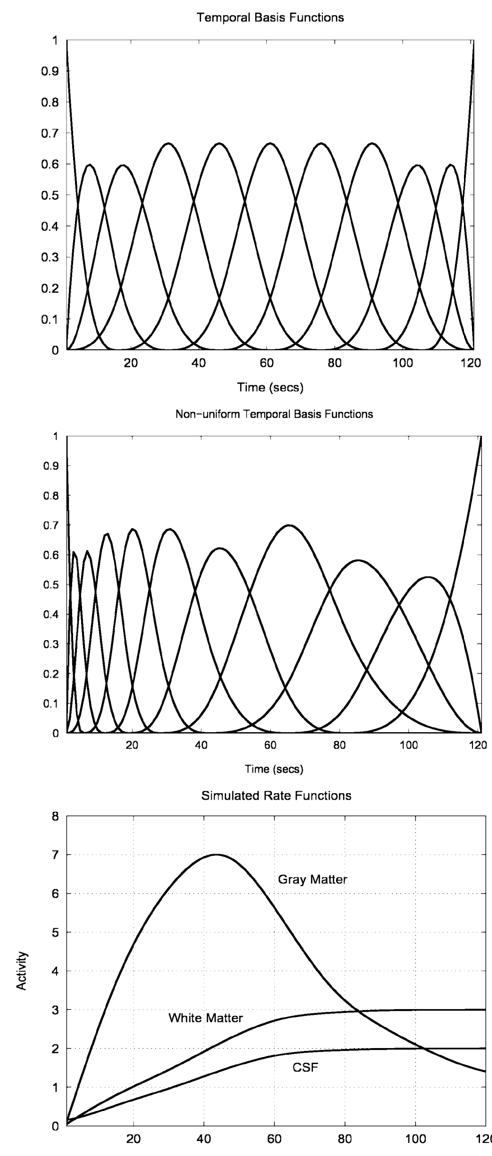

Fig. 4. The set of cubic B-spline temporal basis functions with uniformly spaced (top) and nonuniformly spaced (middle) knots and the simulated rate functions.

A. Mean Approximation

In order to evaluate the accuracy of the mean reconstructed image approximation in (12), we generated 250 datasets with 3M counts over 120 s using 11 uniformly spaced B-splines.

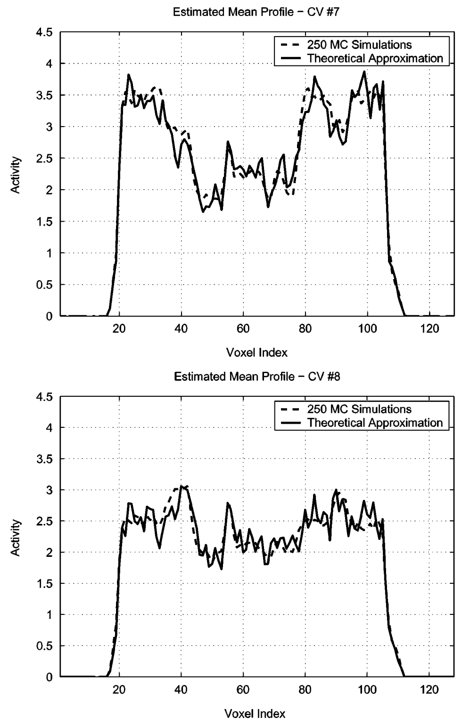

Fig. 5. Central mean reconstructed image estimate profiles (horizontal) for control vertices 7 (top) and 8 (bottom) using Monte Carlo simulations and the theoretical approximation.

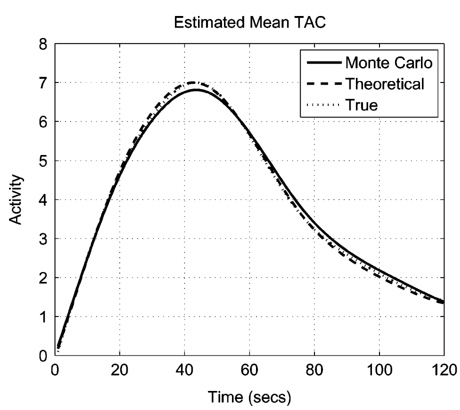

Fig. 6. Estimated mean time activity curves for voxel (60, 70) in gray matter. Solid lines are for the mean approximation in (12) and dashed lines show the mean of all 250 reconstructions. The true activity curve is also shown for reference.

Fig. 5 shows the mean of the Monte Carlo simulations and the mean estimate resulting from (12) for $\beta = 0.01$ and $\gamma = 100$. Our implicitly defined mean estimator provides a mean estimate close to the mean of the 250 Monte Carlo simulations. Fig. 6 shows the estimated mean time activity curves for a voxel in gray matter. The true time activity curve is also shown for reference.

B. Covariance Approximation

In order to validate the covariance approximation in (49), we simulated 500 datasets with 300 K counts each and 200 datasets with 3M counts each using uniformly spaced B-splines. The 300K count datasets simulated medium count scans and 3M count datasets simulated high count scans. We reconstructed the datasets with 300 K counts with $\beta = 0.01$ and $\gamma = 100$. We

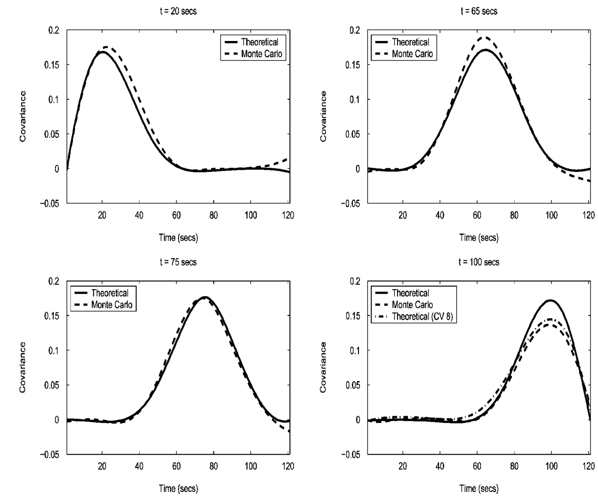

Fig. 7. Covariances between estimates at all times at voxel (64, 64) and at the same voxel at 20 (top-left), 65 (top-right), 75 (bottom-left), and 100 s with 300 K counts, $\beta = 0.01$ $\gamma = 100$.

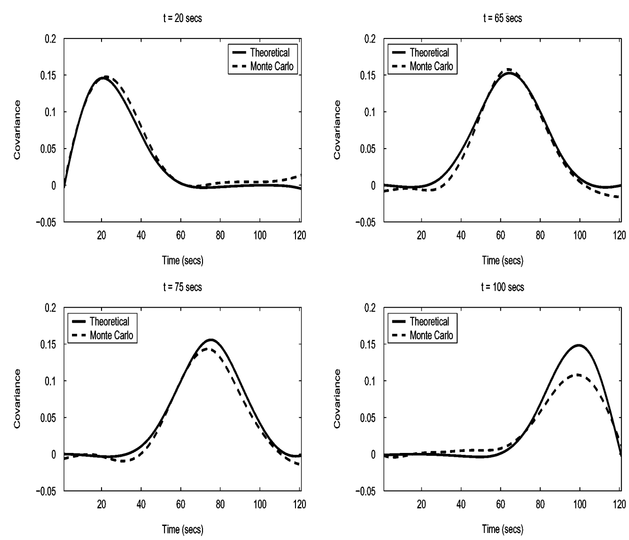

Fig. 8. Covariances between estimates at all times at voxel (63, 64) and at voxel (64, 64) at 20 (top-left), 65 (top-right), 75 (bottom-left), and 100 s with 300 K counts, $\beta = 0.01$ $\gamma = 100$.

used $\beta = 0.1, \gamma = 10$, and $\beta = 0.2, \gamma = 10$ for the datasets with 3M counts. In general, smoothing parameters may have to be varied according to the total number of counts, if approximately uniform resolution over different count rates is desired.

Fig. 7 shows the estimated, by substituting (49)–(50) into (14), and Monte Carlo covariances between activity estimates at all times at voxel (64, 64) and at the same voxel at 20, 65, 75, and 100 s. The theoretical estimates are computed by diagonalizing $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ around their central columns. For $\mathrm{t} = 100$ s, we show an additional theoretical estimate for the case in which the diagonalization is performed around the 8th column (the 8th control vertex is closer to $\mathrm{t} = 100$ s). Theoretically estimated covariances are within 10% of the Monte Carlo simulation covariances.

The covariance between the estimates at voxel (64, 64) at $\mathrm{t} = 20, 65, 75$, and 100 s with the estimates at the neighboring voxel (63, 64) at all time points (i.e., $\mathrm{t} = 0$–120 s) are plotted in Fig. 8. Except for the estimate at $\mathrm{t} = 100$ sec, the theoretical estimates are in good agreement with those obtained from

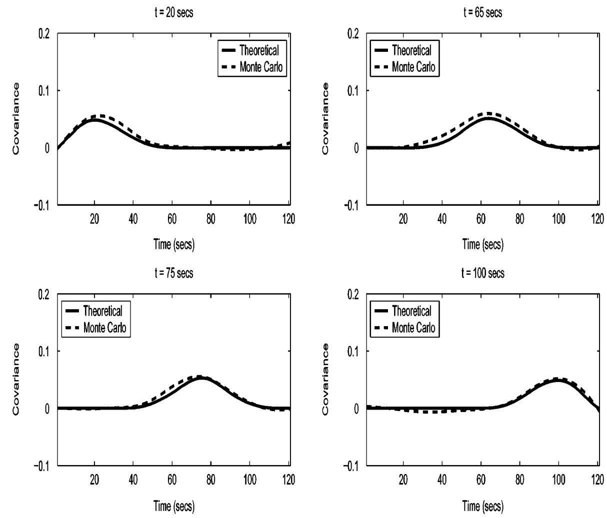

Fig. 9. Covariances between estimates at all times at voxel (64, 64) and at the same voxel at 20 (top-left), 65 (top-right), 75 (bottom-left), and 100 s with 3 M counts, $\beta = 0.2$ $\gamma = 10$.

Monte Carlo simulations. Fig. 9 shows the covariance results at voxel (64, 64) for the second group of simulations with 3M counts/dataset. Again, the theoretical and Monte Carlo covariance estimates are in good agreement with each other. The theoretical approximation accuracy is reduced around $t = 100$ s because the circulant approximations to $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ deteriorate toward the beginning and end of the scan duration even with uniform B-splines as demonstrated in Fig. 1. We also looked at the theoretical and empirical covariance estimates between estimated activity at a white matter voxel and a gray matter voxel at opposite corners of the brain and both showed negligible covariance as expected. Although the estimates are not strictly independent because of contributions to common LORs and the spatial penalty, they are approximately uncorrelated.

We also reconstructed 200 datasets with 3M counts each using the nonuniform B-spline basis shown in Fig. 4. In this case, we ran 100 iterations of conjugate gradient to ensure convergence. The knots were placed at approximately equal arc lengths along the "total activity versus time" curve resulting in denser knots between 10 and 50 s. Fig. 10 shows the estimated and Monte Carlo covariances between activity estimates at all times at voxel (64, 64) and at the same voxel at 10, 20, 40, and 65 s. Fig. 11 shows the covariance between the estimates at voxel (64, 64) at $t = 10, 20, 40$, and 65 s with the estimates at the neighboring voxel (63, 64) at all time points. As shown in Fig. 2, the circulant approximations to $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ are less accurate compared to the uniform B-spline case; however, the theoretical covariance approximations are still in reasonably good agreement with the Monte Carlo simulations. The approximation is more accurate between 15 and 45 s, close to the support of the sixth control vertex (around which the circulant approximations were made) and the approximation accuracy is reduced when moved away from the sixth control vertex.

Note that nonuniform B-splines might be necessary for accurate reconstructions when there are time periods with rapid changes in activity. On the other hand, uniform B-splines result in more accurate theoretical estimates of mean and covariance. Accurate reconstructions and analysis can both be achieved

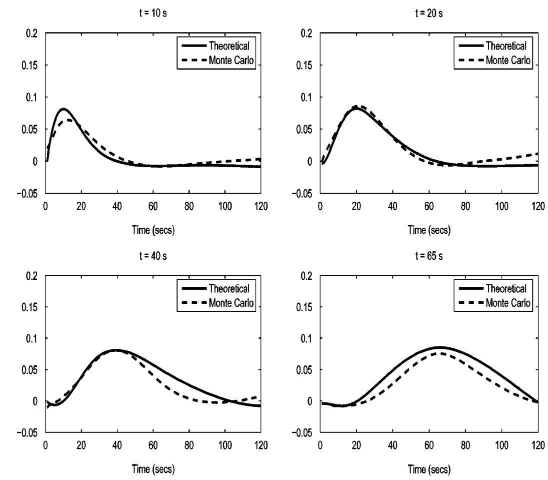

Fig. 10. Covariances between estimates at all times at voxel (64, 64) and at the same voxel at 10 (top-left), 20 (top-right), 40 (bottom-left), and 65 s with 3M counts when nonuniform B-splines are used and $\beta = 0.1$ $\gamma = 100$.

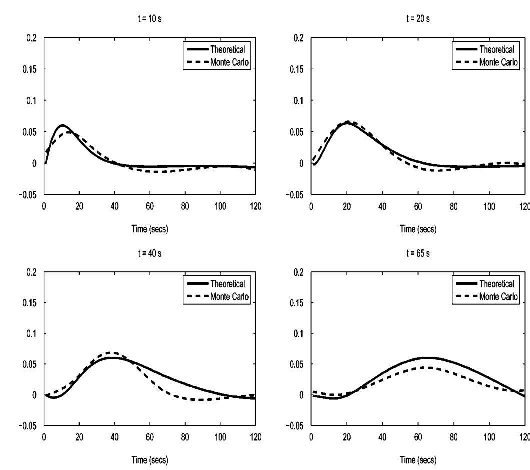

Fig. 11. Covariances between estimates at all times at voxel (64, 64) and at voxel (63, 64) at 10 (top-left), 20 (top-right), 40 (bottom-left), and 65 s with 3M counts when nonuniform B-splines are used and $\beta = 0.01$ $\gamma = 100$.

by warping the time axis so that the "total activity versus time" curve becomes constant and then using uniformly spaced B-splines. Reconstructions and analysis can be performed in the warped-time space with uniform B-splines, and the results mapped back to real time using an inverse warp. This operation is computationally efficient because it requires only a one-to-one transformation on arrival times over a single pass of the list-mode data. An exact homogenization of total activity is not necessary either, it is sufficient that uniform B-splines accurately represent the transformed total activity curve.

C. Dynamic Reconstructions Versus Multiframe Static Reconstructions

An advantage of the theoretical analysis presented in this paper is that it allows for comparisons between different algorithms over a large set of smoothing parameter values at low computational cost. In this subsection, we compare our dynamic penalized ML reconstruction algorithm against the traditional multiframe image reconstruction in terms of the mean and variances of reconstructed time activity curves.

We used the slice of the Hoffman phantom shown in Fig. 3 and used scaled versions of the time activity curves in Fig. 4 for a noiseless sinogram corresponding to a mean of 3M counts over the scan duration of 120 s. The spatial smoothing levels were controlled by spatial hyperparameters (β) in both dynamic and static reconstructions. The degree of temporal smoothing was adjusted by using the temporal hyperparameter (γ) in dynamic reconstructions and by frame durations (for average activity) and the temporal hyperparameter in regularized B-spline fits (for instantaneous activity) in static reconstructions. We used 11 uniformly spaced B-splines in dynamic reconstructions and 12 frames for the static reconstructions. Varying frame durations were used so that there were approximately the same number of counts per frame. We used (12) and (51) for theoretical mean and variance estimation in dynamic PET reconstructions. For multiframe static PET reconstructions, we used the following approximations:

Static Penalized ML [5], [8]

$$E(\hat{\mathbf{x}}) \approx \arg\max_{\mathbf{x}\geq 0} \left(\sum_{i=1}^{n_d} \bar{y}_i \, log \, \bar{y}_i - \bar{y}_i \right) - \beta\phi(\mathbf{x}) \quad (56)$$

$$\mathrm{var}(\hat{x}_{j'}) \approx \frac{\kappa_{j'}^{-2}}{n_v} \sum_{j=1}^{n_v} \frac{\lambda_j^{(s)}}{(\lambda_j^{(s)} + \kappa_{j'}^{-2}\beta\mu_j)^2} \quad (57)$$

Fig. 12 shows the absolute bias versus standard deviation plots for average activity estimates at the gray matter voxel (60, 64) for 8- and 15-s frames centered at $\mathrm{t} = 45$ s. Fig. 13 shows the same absolute bias versus standard deviation curves for the white matter voxel (54, 34) for frames centered at $\mathrm{t} = 60$ s. The static PET curve in each figure is obtained by varying the degree of spatial smoothing. Each dynamic curve corresponds to a fixed value of the temporal smoothing parameter and a bias/variance tradeoff similar to the one in the static case takes place as the spatial smoothing parameter varies. The bias/variance curves corresponding to the dynamic reconstructions lie below and to the left of those corresponding to static reconstructions, indicating that dynamic images have lower variance than multiframe static images at matched bias levels and vice versa. We note that the differences between the bias/variance curves of static and dynamic reconstructions are larger than 10%–15% error margins that we expect in the theoretical approximations.

In order to compare the multiframe and truly dynamic reconstructions in terms of instantaneous activity estimation, we also fit cubic B-splines to the multiframe estimates at each voxel. The fit was performed within a regularized least-squares framework so as to minimize the sum of the least squares error between the multiframe and B-spline average activity estimates and a temporal penalty term. The temporal penalty was the same voxelwise temporal penalty used in dynamic reconstructions. Therefore, the control vertices of the B-spline fit at voxel j are given by

$$\hat{\mathbf{w}}_j^{\mathrm{FIT}} = \arg\min_{\mathbf{w}} \left\| \mathbf{B}_{\mathrm{FIT}}\mathbf{w} - \hat{\mathbf{x}}_j^{\mathrm{STA}} \right\|^2 + \lambda \mathbf{w}^T \mathbf{Z} \mathbf{w} \quad (58)$$

$$\hat{\mathbf{w}}_j^{\mathrm{FIT}} = [\mathbf{B}_{\mathrm{FIT}}^T \mathbf{B}_{\mathrm{FIT}} + \lambda\mathbf{Z}]^{-1} \mathbf{B}_{\mathrm{FIT}}^T \hat{\mathbf{x}}_j^{\mathrm{STA}} \quad (59)$$

where $\hat{\mathbf{w}}_j^{\mathrm{FIT}}$ denotes the coefficients of the cubic B-spline fit at voxel $j, \hat{\mathbf{x}}_j^{\mathrm{STA}}$ contains the activity estimates at voxel j for each frame, λ is the temporal hyperparameter controlling the degree of temporal smoothness, and $\mathbf{B}_{\mathrm{FIT}}$ is a matrix whose ℓkth element is given by

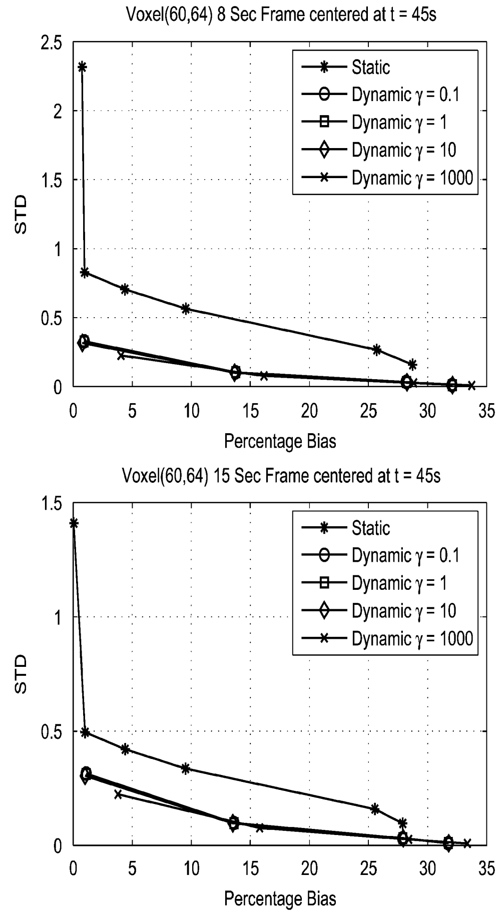

Fig. 12. Percentage bias versus standard deviation plots for multiframe static and dynamic penalized ML reconstructions with various degrees of spatial and temporal smoothing (varies only in dynamic reconstructions) for average activity estimates over 8- (top) and 15-s frames centered at $\mathrm{t} = 45$ s for the gray matter voxel (60, 64).

$$[\mathbf{B}_{\mathrm{FIT}}]_{\ell k} = \int_{t_{k-1}}^{t_k} B_\ell(t)\, dt \quad (60)$$

where t_{k-1} and t_k denote the end points of frame k. Since the coefficients of the B-spline fit are obtained by performing a linear operation on independently reconstructed multiframe estimates, their mean and variance properties follow directly from those of multiframe estimates

$$E\left\{\hat{\mathbf{w}}_j^{\mathrm{FIT}}\right\} = \left[\mathbf{B}_{\mathrm{FIT}}^T\mathbf{B}_{\mathrm{FIT}} + \lambda\mathbf{Z}\right]^{-1} \mathbf{B}_{\mathrm{FIT}}^T E\left\{\hat{\mathbf{x}}_j^{\mathrm{STA}}\right\} \quad (61)$$

$$\mathrm{cov}\left(\hat{\mathbf{w}}_j^{\mathrm{FIT}}\right) = \left[\mathbf{B}_{\mathrm{FIT}}^T\mathbf{B}_{\mathrm{FIT}} + \lambda\mathbf{Z}\right]^{-1} \mathbf{B}_{\mathrm{FIT}}^T \mathbf{D}\left\{\mathrm{var}\left(\hat{x}_j^f\right)\right\} \times \mathbf{B}_{\mathrm{FIT}} \left[\mathbf{B}_{\mathrm{FIT}}^T\mathbf{B}_{\mathrm{FIT}} + \lambda\mathbf{Z}\right]^{-T} \quad (62)$$

where $\hat{x}_j^f$ denotes the static activity estimate for voxel j, frame f. The comparison between the regularized B-spline fits and dynamic reconstructions are shown in Figs. 14 and 15. The bias/variance curves for dynamic reconstructions lie below those for the multiframe reconstructions indicating more accurate estimates can be obtained using the dynamic list-mode approach rather than the two-step procedure of multiframe activity estimation followed by a B-spline fit.

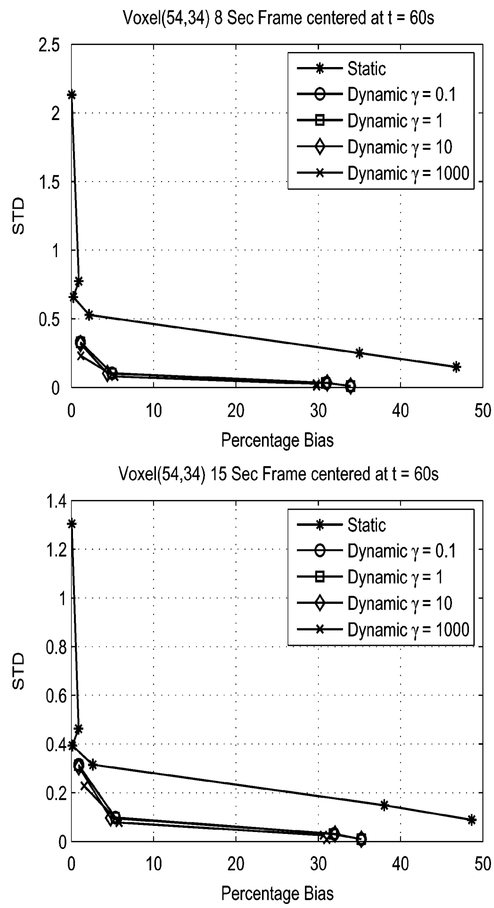

Fig. 13. Percentage bias versus standard deviation plots for multiframe static and dynamic penalized ML reconstructions with various degrees of spatial and temporal smoothing (varies only in dynamic reconstructions) for average activity estimates over 8- (top) and 15-s frames centered at $t = 60$ s for the white matter voxel (54, 34).

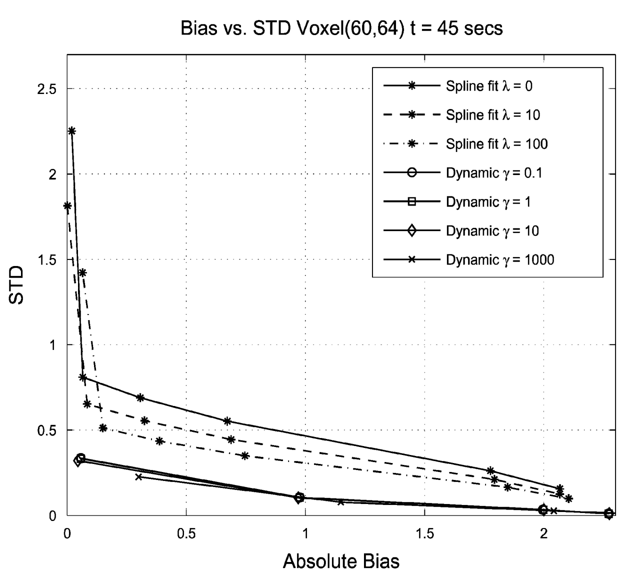

Fig. 14. Absolute bias versus standard deviation plots for B-spline fits to multiframe static and dynamic penalized ML reconstructions for various degrees of spatial and temporal smoothing for activity estimates at the white matter voxel (60, 64) at $t = 45$ s.

We also note that, although dynamic reconstructions provide more accurate time-activity curve estimates, if we were to subsequently fit parametric models to the data, it does not necessarily follow that the resulting parameter estimates will be better using our approach. The goodness of the fits would depend on the particular kinetic model, degree of smoothing and the temporal basis functions [17].

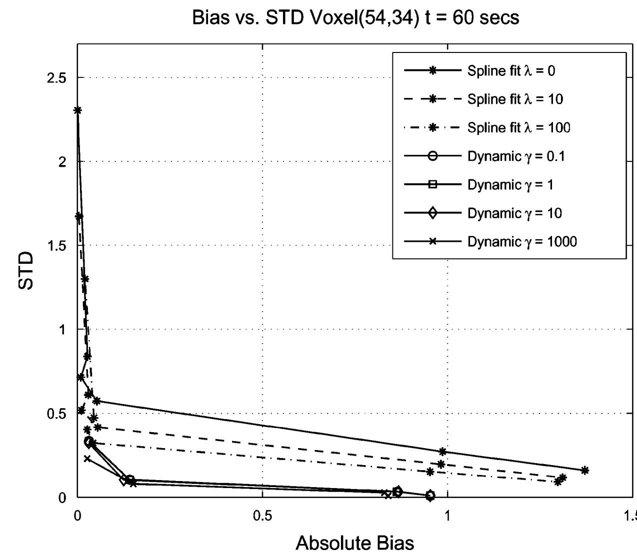

Fig. 15. Absolute bias versus standard deviation plots for B-spline fits to multiframe static and dynamic penalized ML reconstructions for various degrees of spatial and temporal smoothing for activity estimates at the white matter voxel (54, 34) at $t = 60$ s.

IV. Conclusion

In order to evaluate the performance of our continuous time dynamic PET reconstruction algorithm [16], we derived procedures for approximating means and variances of dynamic average and instantaneous rate estimates. We used DFT-based diagonalizations as in [8] and [22] to perform the matrix inversions which would otherwise not be feasible except for 1-D cases. This performance analysis allows us to evaluate the performances of all dynamic reconstructions in which the rate functions are parameterized by continuous basis functions and penalty terms are quadratic and separable in control vertices. Although we used cubic B-splines in our work, the analysis is applicable to all differentiable temporal basis functions.

The approximations were compared with Monte Carlo simulations revealing generally good agreement. However, errors increased, toward the endpoints of the time series, which correspond to the locations at which the circulant approximations are least accurate. These errors were reduced by diagonalizing the temporal matrices $\mathbf{B}^T\mathbf{B}$ and $\mathbf{T}$ about the control vertex that best spans the time interval of interest. Similarly, accuracy also decreased when using nonuniformly spaced B-splines. In this case, we could circumvent the problem by using a nonlinear warping in time in combination with uniformly spaced B-splines.

We also used the analytic approximations to compare our penalized ML dynamic reconstruction algorithm against a multiframe static penalized ML approach over a large range of smoothing parameters without extensive Monte Carlo simulations. The resulting bias versus variance curves indicated that our dynamic reconstruction algorithm could provide lower variance at matched bias levels and *vice versa* compared to the multiframe approach.

APPENDIX

In this appendix we show that the maximizer of the bin-mode log-likelihood function can be made *arbitrarily close* to the maximizer of the continuous-time objective function by choosing sufficiently small bin-widths under the following assumptions.

1) The continuous-time objective function is strictly concave and all constraints are inactive at its maximum.
2) The maximizer of the continuous-time objective function is bounded.
3) Temporal basis functions have bounded derivatives.

Lemma 1: If $\nabla^2\Phi_{\mathrm{CT}}(\mathbf{w}) < 0$ and $|B'_\ell(t)| \leq C < \infty \, \forall \, \ell, t$, then there exists a sufficiently small δ where $\max_{i,n} \Delta t_{\mathrm{in}} \leq \delta$ such that $\mathbf{w}^*_{\mathrm{CT}} = \arg\max_{\mathbf{w}\in(\mathcal{W}\cap\mathcal{B})} \Phi_{\mathrm{CT}}(\mathbf{w})$ and $\mathbf{w}^*_{\mathrm{BM},N} = \arg\max_{\mathbf{w}\in(\mathcal{W}\cap\mathcal{B})} \Phi^{(N)}_{\mathrm{BM}}(\mathbf{w})$ satisfy

$$\|\mathbf{w}^*_{\mathrm{BM},N} - \mathbf{w}^*_{\mathrm{CT}}\|^2 \leq K\delta \tag{63}$$

where K is a constant, $\Phi^{(N)}_{\mathrm{BM}}(\mathbf{w})$ is the bin-mode objective function with the scan duration divided into N bins, $\Delta t_{\mathrm{in}} \equiv t_{\mathrm{in}} - t_{i(n-1)}$ is the nth temporal bin-width at the ith detector pair, and $\max_{i,n} \Delta t_{\mathrm{in}}$ is sufficiently small so that there is at most one event per bin. Here, we present the proof for the more general case when bin-widths and boundaries can vary at each detector pair, however, the proof for the constant, uniform bin-width case is sufficient for our covariance approximations. The sets $\mathcal{W}$ and $\mathcal{B}$ are given by

$$\mathcal{W} \equiv \left\{ \mathbf{w} \left| \sum_{\ell=1}^{n_b} w_{j\ell} B_\ell(t) \geq 0 \forall j, t; \right. \right.$$
$$\left. \sum_{j=1}^{n_v}\sum_{\ell=1}^{n_b} p_{ij} w_{j\ell} B_\ell(a_{ik}) > 0 \, \forall \, i, k \right\} \tag{64}$$

$$\mathcal{B} \equiv \{\mathbf{w} | -U \leq w_{j\ell} \leq U, \forall j, \ell\} \tag{65}$$

where U is a sufficiently large constant such that $\mathbf{w}^*_{\mathrm{CT}} \in \mathcal{B}$.

We start our proof with the bin-mode objective function that results from dividing the scan duration into N (possibly nonuniform) bins

$$L^{(N)}_{\mathrm{BM}}(\mathbf{w}) = \sum_{i=1}^{n_d}\sum_{n=1}^{N} y_i^{(n)} \log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} \int_{t_{i(n-1)}}^{t_{\mathrm{in}}} B_\ell(t)\, dt\right)$$
$$- \sum_{i=1}^{n_d}\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} \int_0^T B_\ell(t)\, dt. \tag{66}$$

Since there can not be any two events at *exactly* the same time under the inhomogeneous Poisson process model, $\max_{i,n} \Delta t_{\mathrm{in}}$ can always be chosen small enough such that there is *at most* one event per bin. Therefore, the bin mode log-likelihood can be written as

$$L^{(N)}_{\mathrm{BM}}(\mathbf{w}) = \sum_{(i,n)\in\mathcal{E}} y_i^{(n)} \log\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} \int_{t_{i(n-1)}}^{t_{\mathrm{in}}} B_\ell(t)\, dt\right)$$
$$- \sum_{i=1}^{n_d}\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} \int_0^T B_\ell(t)\, dt \tag{67}$$

where $\mathcal{E} \equiv \{(i,n) | y_i^{(n)} = 1\}$ is the set of spatiotemporal bins that recorded an event. We now define $\eta_i(t)$ as the rate function at the ith detector pair

$$\eta_i(t) \equiv \sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} B_\ell(t) \tag{68}$$

and use the mean value theorem at each $(i,n) \in \mathcal{E}$ to write

$$\int_{t_{i(n-1)}}^{t_{\mathrm{in}}} \eta_i(t)\, dt = \eta_i(\tilde{c}_{\mathrm{in}})\Delta t_{\mathrm{in}} \tag{69}$$

where $t_{i(n-1)} \leq \tilde{c}_{\mathrm{in}} \leq t_{\mathrm{in}}$. Substituting this into (66) we get

$$\begin{aligned} L^{(N)}_{\mathrm{BM}}(\mathbf{w}) &= \sum_{(i,n)\in\mathcal{E}} y_i^{(n)} \log(\eta_i(\tilde{c}_{\mathrm{in}})\Delta t_{\mathrm{in}}) \\ &\quad - \sum_{i=1}^{n_d}\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} \int_0^T B_\ell(t)\, dt \\ &= \sum_{(i,n)\in\mathcal{E}} \log(\eta_i(\tilde{c}_{\mathrm{in}})) - \sum_{i=1}^{n_d}\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} \int_0^T B_\ell(t)\, dt \\ &\quad + \sum_{(i,n)\in\mathcal{E}} \log \Delta t_{\mathrm{in}}. \end{aligned} \tag{70}$$

Since the last term is independent of $\mathbf{w}$, we can define a new objective function that has the same maximizer as $L_{\mathrm{BM}}(\mathbf{w})$

$$\tilde{L}_{\mathrm{BM}}(\mathbf{w}) = \sum_{(i,n)\in\mathcal{E}} \log(\eta_i(\tilde{c}_{\mathrm{in}}))$$
$$- \sum_{i=1}^{n_d}\left(\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} \int_0^T B_\ell(t)\, dt\right). \tag{71}$$

The norm of the difference between $\tilde{L}_{\mathrm{BM}}(\mathbf{w})$ and $L_{\mathrm{CT}}(\mathbf{w})$ is given by [see (1)]

$$\|\tilde{L}_{\mathrm{BM}}(\mathbf{w}) - L_{\mathrm{CT}}(\mathbf{w})\| = \left\| \sum_{(i,n)\in\mathcal{E}} \log\left(\frac{\eta_i(\tilde{c}_{\mathrm{in}})}{\eta_i(a_{\mathrm{in}})}\right)\right\|. \tag{72}$$

Note that $\eta_i(t)$ is a continuous function of t and $\eta_i(a_{ik}) > 0$ for all i, k for all $\mathbf{w} \in \mathcal{W}$. Since $\eta_i(t)$ is continuous, there exists an interval $\mathcal{I}_{ik}$ around a_{ik} where $\eta_i(t) > 0$ holds for $t \in \mathcal{I}_{ik}$. Because of this property, $\log \eta_i(t)$ has a bounded derivative and is, therefore, Lipschitz continuous at $\mathcal{I}_{ik}$ for all $i, k, \mathbf{w} \in \mathcal{W}$

$$\left\|\frac{\partial \log \eta_i(t)}{\partial t}\right\| = \left\|\frac{\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} B'_\ell(t)}{\sum_{\ell=1}^{n_b}\sum_{j=1}^{n_v} p_{ij} w_{j\ell} B_\ell(t)}\right\| \leq C < \infty. \tag{73}$$

The numerator is bounded due to the boundedness of the derivatives of the temporal basis functions and the boundedness of $\mathcal{W} \cap \mathcal{B}$ and the denominator is strictly positive over all $\mathcal{I}_{ik}$ for all $\mathbf{w} \in \mathcal{W}$ because of the continuity of $\eta_i(t)$, hence $C < \infty$.

Note that we have $t_{i(n-1)} \leq \tilde{c}_{\mathrm{in}} \leq t_{\mathrm{in}}$ due to the mean value theorem and $t_{i(n-1)} \leq a_{\mathrm{in}} \leq t_{\mathrm{in}}$ holds by definition. Therefore, $\max_{i,n} \Delta t_{\mathrm{in}} \leq \delta$ implies $\|\tilde{c}_{\mathrm{in}} - a_{\mathrm{in}}\| \leq \delta$ and by choosing $\delta \leq \min_{ik} |\mathbf{I}_{ik}|/2$ we ensure the Lipschitz continuity of $\eta_i(t)$ over $\mathcal{I}_{ik}$ for all i, k to get

$$\|\eta_i(\tilde{c}_{\mathrm{in}}) - \eta_i(a_{\mathrm{in}})\| \leq L\|\tilde{c}_{\mathrm{in}} - a_{\mathrm{in}}\|. \tag{74}$$

Using the definition of δ we have

$$\left\| \sum_{(i,n)\in\mathcal{E}} log\left(\frac{\eta_i(\tilde{c}_{\mathrm{in}})}{\eta_i(a_{\mathrm{in}})}\right)\right\| \leq LY\delta \tag{75}$$

where $Y = \sum_{i=1}^{n_d} x_i$, denotes the total number of events in the scan. Since the penalty terms are identical in both bin-width and continuous-time objective functions, we also have $\|\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}) - \Phi_{\mathrm{CT}}(\mathbf{w})\| \leq LY\delta$ for all $\mathbf{w} \in (\mathcal{W} \cap \mathcal{B})$ where

$$\Phi_{\mathrm{BM}}^{(N)}(\mathbf{w}) = L_{\mathrm{BM}}^{(N)}(\mathbf{w}) - \frac{1}{2}\mathbf{w}^T(\beta\mathbf{R}_S + \gamma\mathbf{R}_T)\mathbf{w} \tag{76}$$

$$\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}) = \tilde{L}_{\mathrm{BM}}^{(N)}(\mathbf{w}) - \frac{1}{2}\mathbf{w}^T(\beta\mathbf{R}_S + \gamma\mathbf{R}_T)\mathbf{w} \tag{77}$$

$$\Phi_{\mathrm{CT}}(\mathbf{w}) = L_{\mathrm{CT}}(\mathbf{w}) - \frac{1}{2}\mathbf{w}^T(\beta\mathbf{R}_S + \gamma\mathbf{R}_T)\mathbf{w} \tag{78}$$

Until this point we showed that if $\max_{i,n}\Delta t_{\mathrm{in}} \leq \delta$ then $\|\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}) - \Phi_{\mathrm{CT}}(\mathbf{w})\| \leq LT\delta \equiv \epsilon$ for sufficiently small $\delta > 0$. Next, we use this property to show that if $\Phi_{\mathrm{CT}}(\mathbf{w}_1) - \Phi_{\mathrm{CT}}(\mathbf{w}_2) > 2\epsilon$ holds then $\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_1) > \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_2)$ for any two vectors $\mathbf{w}_1$ and $\mathbf{w}_2$

$$\Phi_{\mathrm{CT}}(\mathbf{w}_1) - \Phi_{\mathrm{CT}}(\mathbf{w}_2) > 2\epsilon \tag{79}$$

$$\left\|\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_1) - \Phi_{\mathrm{CT}}(\mathbf{w}_1)\right\| \leq \epsilon$$
$$\Rightarrow \Phi_{\mathrm{CT}}(\mathbf{w}_1) - \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_1) \leq \epsilon \tag{80}$$

$$\left\|\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_2) - \Phi_{\mathrm{CT}}(\mathbf{w}_2)\right\| \leq \epsilon$$
$$\Rightarrow \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_2) - \Phi_{\mathrm{CT}}(\mathbf{w}_2) \leq \epsilon \tag{81}$$

$$\Phi_{\mathrm{CT}}(\mathbf{w}_1) - \Phi_{\mathrm{CT}}(\mathbf{w}_2) - 2\epsilon$$
$$\leq \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_1) - \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_2) \tag{82}$$

$$\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_2) < \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_1). \tag{83}$$

This also implies that if $\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_2) \geq \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_1)$, then $\Phi_{\mathrm{CT}}(\mathbf{w}_1) - \Phi_{\mathrm{CT}}(\mathbf{w}_2) \leq 2\epsilon$. Since $\mathbf{w}_{\mathrm{BM},N}^*$ is the global maximum of $\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w})$ we have $\tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_{\mathrm{BM},N}^*) \geq \tilde{\Phi}_{\mathrm{BM}}^{(N)}(\mathbf{w}_{\mathrm{CT}}^*)$ and (79)–(83) imply

$$\Phi_{\mathrm{CT}}(\mathbf{w}_{\mathrm{CT}}^*) - \Phi_{\mathrm{CT}}(\mathbf{w}_{\mathrm{BM},N}^*) \leq 2\epsilon. \tag{84}$$

What we have just shown is that if the bin-mode objective function differs from the continuous-time objective function by only ϵ at any vector inside the feasible set, then their maximum values can not differ by more than 2ϵ. Finally, we use the Taylor series expansion (with remainder) of $\Phi_{\mathrm{CT}}(\mathbf{w})$ at $\mathbf{w}_{\mathrm{CT}}^*$ evaluated at $\mathbf{w}_{\mathrm{BM},N}^*$

$$\Phi_{\mathrm{CT}}(\mathbf{w}_{\mathrm{BM},N}^*) = \Phi_{\mathrm{CT}}(\mathbf{w}_{\mathrm{CT}}^*)$$
$$+(\mathbf{w}_{\mathrm{BM},N}^* - \mathbf{w}_{\mathrm{CT}}^*)^T\nabla^2 L_{\mathrm{CT}}(\tilde{\mathbf{w}})(\mathbf{w}_{\mathrm{BM},N}^* - \mathbf{w}_{\mathrm{CT}}^*) \tag{85}$$

where $\tilde{\mathbf{w}}$ is between $\mathbf{w}_{\mathrm{BM},N}^*$ and $\mathbf{w}_{\mathrm{CT}}^*$. Note that $\nabla L_{\mathrm{CT}}(\mathbf{w}_{\mathrm{CT}}^*) = \mathbf{0}$. Since $-\nabla^2 L_{\mathrm{CT}}(\tilde{\mathbf{w}})$ is positive definite, we have

$$\lambda_{\min}^2\|\mathbf{w}_{\mathrm{BM},N}^* - \mathbf{w}_{\mathrm{CT}}^*\|^2$$
$$\leq -(\mathbf{w}_{\mathrm{BM},N}^* - \mathbf{w}_{\mathrm{CT}}^*)^T\nabla^2 L_{\mathrm{CT}}(\tilde{\mathbf{w}})(\mathbf{w}_{\mathrm{BM},N}^* - \mathbf{w}_{\mathrm{CT}}^*)$$
$$\leq 2\epsilon \tag{86}$$

where $\lambda_{\min} > 0$ denotes the minimum eigenvalue of $-\nabla^2 L_{\mathrm{CT}}(\tilde{\mathbf{w}})$. Hence

$$\|\mathbf{w}_{\mathrm{BM},N}^* - \mathbf{w}_{\mathrm{CT}}^*\|^2 \leq \frac{2LY}{\lambda_{\min}^2}\delta \equiv K\delta. \tag{87}$$

References

[1] H. H. Barrett, D. W. Wilson, and B. M. W. Tsui, "Noise properties of the EM algorithm. I. Theory," *Phys. Med. Biol.*, vol. 39, no. 5, pp. 833–846, May 1994.

[2] D. W. Wilson, B. M. W. Tsui, and H. H. Barrett, "Noise properties of the EM algorithm. II. Monte Carlo simulations," *Phys. Med. Biol.*, vol. 39, no. 5, pp. 847–871, May 1994.

[3] W. Wang and G. Gindi, "Noise analysis of MAP-EM algorithms for emission tomography," *Phys. Med. Biol.*, vol. 42, no. 11, pp. 2215–2232, Nov. 1997.

[4] H. Erdoğan and J. A. Fessler, "Ordered subsets algorithms for transmission tomography," *Phys. Med. Biol.*, vol. 44, pp. 2835–2851, Jul. 1999.

[5] J. A. Fessler, "Mean and variance of implicitly defined biased estimators (such as penalized maximum likelihood): Applications to tomography," *IEEE Trans. Image Process.*, vol. 5, no. 3, pp. 493–506, Mar. 1996.

[6] J. A. Fessler and W. L. Rogers, "Spatial resolution properties of penalized-likelihood image reconstruction: Space-invariant tomographs," *IEEE Trans. Image Process.*, vol. 5, no. 9, pp. 1346–1358, Sep. 1996.

[7] J. Qi and R. M. Leahy, "A theoretical study of the contrast recovery and variance of MAP reconstructions from PET data," *IEEE Trans. Med. Imag.*, vol. 18, no. 4, pp. 293–305, Apr. 1999.

[8] ——, "Resolution and noise properties of MAP reconstruction for fully 3-D PET," *IEEE Trans. Med. Imag.*, vol. 19, no. 5, pp. 493–506, May 2000.

[9] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*. Orlando, FL: Academic, 1985.

[10] C. K. Abbey, "Assessment of Reconstructed Images," Ph.D. thesis, Univ. Arizona, Tucson, 1998.

[11] J. Qi, "A unified noise analysis for iterative image estimation," *Phys. Med. Biol.*, vol. 48, no. 21, pp. 3505–3519, Nov. 2003.

[12] J. W. Stayman and J. A. Fessler, "Regularization for uniform spatial resolution properties in penalized-likelihood image reconstruction," *IEEE Trans. Med. Imag.*, vol. 19, no. 6, pp. 601–615, Jun. 2000.

[13] ——, "Nonnegative definite quadratic penalty design for penalized-likelihood reconstruction," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 2, San Diego, CA, 2001, pp. 1060–1063.

[14] S. Mustafovic, K. Thielemans, D. Hogg, and P. Bloomfield, "Object dependency of resolution and convergence rate in OSEM with filtering," in *Proc. IEEE Nuclear Science Symp. Medical Imaging Conf.*, vol. 3, San Diego, CA, 2001, pp. 1786–1790.

[15] S. Mustafovic and K. Thielemans, "Object dependency of resolution in reconstruction algorithms with interiteration filtering applied to PET data," *IEEE Trans. Med. Imag.*, vol. 23, no. 4, pp. 433–446, Apr. 2004.

[16] T. E. Nichols, J. Qi, E. Asma, and R. M. Leahy, "Spatiotemporal reconstruction of list-mode PET data," *IEEE Trans. Med. Imag.*, vol. 21, no. 4, pp. 396–405, Apr. 2002.

[17] S. Ahn, J. A. Fessler, T. E. Nichols, and R. A. Koeppe, "Covariance of kinetic parameter estimators based on time activity curve reconstructions: Preliminary study on 1-D dynamic imaging," presented at the Int. Symp. Biomedical Imaging, Arlington, VA, 2004.

[18] J. Nuyts and J. A. Fessler, "A penalized-likelihood image reconstruction method for emission tomography, compared to postsmoothed maximum-likelihood with matched spatial resolution," *IEEE Trans. Med. Imag.*, vol. 22, no. 9, pp. 1042–1052, Sep. 2003.

[19] Q. Li, E. Asma, and R. M. Leahy, "A fast fully 4D incremental gradient reconstruction algorithm for list mode PET data," presented at the Int. Symp. Biomedical Imaging, Arlington, VA, 2004.

[20] Q. Li, E. Asma, J. Qi, J. R. Bading, and R. M. Leahy, "Accurate estimation of the fisher information matrix for the PET image reconstruction problem," *IEEE Trans. Med. Imag.*, vol. 23, no. 9, pp. 1057–1064, Sep. 2004.

[21] C. de Boor, *A Practical Guide to Splines*, ser. Applied Mathematical Sciences. New York: Springer, 1978, vol. 27 .

[22] J. A. Fessler and S. D. Booth, "Conjugate-gradient preconditioning methods for shift-variant PET image reconstruction," *IEEE Trans. Image Process.*, vol. 8, no. 5, pp. 688–699, May 1999.

[23] A. Rosenfeld and A. Kak, *Digital Picture Processing*, 2nd ed. New York: Academic, 1982, vol. 1.