Published in final edited form as:

IEEE Trans Med Imaging. 2014 January ; 33(1): 173–185. doi:10.1109/TMI.2013.2283229.

Sparsity Constrained Mixture Modeling for the Estimation of Kinetic Parameters in Dynamic PET

Yanguang Lin [Student Member, IEEE],

Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA

Justin P. Haldar [Member, IEEE],

Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA

Quanzheng Li [Member, IEEE],

Department of Radiology, Massachusetts General Hospital, Boston, MA 02114 USA

Peter S. Conti, and

Department of Radiology, University of Southern California, Los Angeles, CA 90033 USA

Richard M. Leahy [Fellow, IEEE]

Signal and Image Processing Institute, University of Southern California, Los Angeles, CA 90089 USA

Abstract

The estimation and analysis of kinetic parameters in dynamic PET is frequently confounded by tissue heterogeneity and partial volume effects. We propose a new constrained model of dynamic PET to address these limitations. The proposed formulation incorporates an explicit mixture model in which each image voxel is represented as a mixture of different pure tissue types with distinct temporal dynamics. We use Cramér-Rao lower bounds to demonstrate that the use of prior information is important to stabilize parameter estimation with this model. As a result, we propose a constrained formulation of the estimation problem that we solve using a two-stage algorithm. In the first stage, a sparse signal processing method is applied to estimate the rate parameters for the different tissue compartments from the noisy PET time series. In the second stage, tissue fractions and the linear parameters of different time activity curves are estimated using a combination of spatial-regularity and fractional mixture constraints. A block coordinate descent algorithm is combined with a manifold search to robustly estimate these parameters. The method is evaluated with both simulated and experimental dynamic PET data.

Keywords

Dynamic PET; Compartment Model; Mixture Modeling; Sparsity; Kinetic Parameter Estimation

I. INTRODUCTION

Dynamic PET involves imaging the temporal dynamics of the spatial distribution of a radioactive tracer, which provides information about molecular processes in the living body.

Compared to conventional static PET, which reconstructs a single image representing the time-averaged spatial tracer distribution, the additional temporal information in dynamic PET provides new capabilities for identifying and differentiating tissues with different metabolic characteristics. However, while this additional dimension of information can be extremely useful, the signal-to-noise ratio (SNR) per time point is very low compared to static PET, and novel analysis techniques are necessary to extract practically-useful information from the time series of images.

Parametric pharmacokinetic models [1], [2] are commonly used to analyze PET data, and provide physiologically-meaningful kinetic parameter estimates that can play an important role in disease diagnosis and treatment evaluation [3], [4]. To compensate for the low SNR in the reconstructed images of dynamic PET, conventional methods used for kinetic parameter estimation often define a region of interest (ROI) and then perform the estimation based on the ROI average [5], [6], [7], [8]. These approaches make the assumption that the ensemble-average signal from the ROI provides a sufficient summary, and that spatial variations in tracer kinetics within the ROI (due to, e.g., partial volume artifacts and tissue heterogeneity) are less important and can be neglected.

However, there are many situations where neglecting spatial heterogeneity might be undesirable. For example, biological heterogeneity is a common feature of malignant tumors. This aspect of malignancy has long been known and classically described with histological features and physiological characteristics. Highly heterogeneous tumors are considered more aggressive with a higher propensity for metastasis or invasion [9]. Therefore quantification of tumor heterogeneity could prove to be a useful metric for treatment assessment or a predictive indicator for treatment failure [10]. As a result, there is a need for pharmacokinetic modeling that preserves information about the spatial heterogeneity of the dynamic PET signal.

Some studies have tried to avoid the homogeneous ROI assumption by estimating kinetic parameters voxel by voxel, resulting in a set of parametric images. Since the PET images have low SNR, existing approaches have used additional constraints to stabilize parameter estimation. Examples include the use of spatial smoothness regularization on the estimated parameter images [11], [12], the use of Tikhonov regularization to ensure that the estimated parameters are close to physiologically reasonable values [13], and the use of regularization to encourage the reconstructed parametric images to have high mutual information with reference anatomical MRI images [14].

The methods described above assume that there is a single kinetic process per voxel, and can work reasonably well when this assumption is valid. However, biological heterogeneity, low spatial resolution, and partial volume effects frequently mean that the signal from each voxel represents a mixture of multiple kinetic processes with distinct dynamic characteristics. As a result, the use of a single kinetic process per voxel can limit performance, particularly in the context of small or highly heterogeneous tumors.

To better model tissue heterogeneity, mixture models have been proposed to model the TAC from each voxel as a linear combination of several basis. Among them, various nonpara-

metric approaches for time series analysis have been applied to analyze dynamic PET data. In [15], [16], principle component analysis (PCA) was used to identify region of interests (ROIs) and study the underlying components of dynamic PET data. In [17], [18], independent component analysis was used to extract a blood input function and other functional components from dynamic PET data. In [19], [20], nonnegative matrix factorization and factor analysis were used to study cardiac dynamic PET data with the incorporation of nonnegativity constraints on the underlying components.

Although effective in certain applications, the nonparametric mixture model methods mentioned above do not make use of the prior information that the dynamic PET signal is constrained by the physics and physiology of metabolism. As a result, the estimated components or factors are unlikely to match closely with physical and physiological expectations, especially when the model fitting is ill-conditioned.

To overcome these limitations, parametric mixture models [21] have been proposed that model the TAC from each voxel as a linear combination of several TACs representing the tracer dynamics of pure tissues. But since parametric mixture models will have more parameters to estimate than single-process parametric models, they generally require the use of more prior information to ensure unique and stable solutions to the parameter estimation problem. Various kinds of prior information have been used to stabilize the estimation of parametric PET mixture models. In [22], [23], basis pursuit and sparse Bayesian learning approaches were proposed to estimate tissue kinetic parameters, under the assumption that the number of distinct tissue types contributing to any given voxel is relatively small. These methods operated voxel-by-voxel, without incorporating the prior information that PET tracer kinetics are often spatially-correlated and smoothly varying. In [24], spatial smoothness constraints were combined with an assumption that the total number of distinct mixture components contributing to the entire PET dataset was small. The spatial smoothness constraint is imposed by explicitly modeling the spatial covariance with a conditional autoregressive model. In [25], a locally constrained mixture model was proposed that allows at most two pure mixture components to be present in any voxel, but with a potentially large number of mixture components present within the entire PET dataset.

In this work, we introduce a new parametric mixture model for dynamic PET that leverages the complementary advantages of some of these different models and constraints while mitigating their disadvantages. In addition, motivated by theoretical analysis of the inverse problem, we introduce a novel sparsity constraint to further improve the stability of the model fitting problem. A two-stage algorithm, which separates the estimation of nonlinear rate parameters and linear terms in the model, is proposed to robustly solve the resulting problem. In the first stage, we estimate the number of tissue types and the nonlinear terms of the tissue kinetic model using the prior information that the overall number of tissue types is small and the tissue fractions for each voxel have sparse structure. In the second stage, we estimate the tissue fractions and the linear terms of the tissue kinetic model using the prior information that the fractions are spatially smooth and the metabolically inactive tissue fraction is sparse. A preliminary version of this work was first described in [26].

The paper is organized in the following way. In section II, we describe the pharmacokinetic and mixture models we use in this work. In section III, we perform a Cramér-Rao lower bound analysis of the mixture model, and observe that the use of prior information can significantly improve the stability of parameter estimation with this model. In section IV, we propose and describe a new approach to constrained estimation of the mixture model parameters. In section V, simulation and real data results are shown to demonstrate and evaluate the capabilities of the proposed approach. Our conclusions are presented in section VII.

II. Modeling Assumptions

A. Pharmacokinetic Model

This work assumes that tracer pharmacokinetics follow a compartment model [27] that describes the PET tracer as it interacts with biological tissues and transitions between different states. Specifically, we use the three-compartment model [28] (shown in Fig. 1) that is commonly used in dynamic F^{18} -Fluorodeoxyglucose (FDG) PET studies. FDG is an analog of glucose whose uptake reflects glucose metabolic rate. The three compartments of the model are: $C_p(t)$, the tracer activity in the blood input function, $C_f(t)$, the tracer activity of the free tracer in tissue, and $C_b(t)$, the activity in partially metabolized FDG. FDG is phosphorylated in the same way as glucose, but cannot be further metabolized and thus accumulates in cells that consume glucose. Kinetics within a homogeneous ROI are governed by rate parameters (k_1 , k_2 , k_3 and k_4) and the following ordinary differential equations:

$$\frac{dC_f(t)}{dt} = k_1 C_p(t) + k_4 C_b(t) - (k_2 + k_3 + \lambda_d) C_f(t) \quad (1)$$

$$\frac{dC_b(t)}{dt} = k_3 C_f(t) - (k_4 + \lambda_d) C_b(t), \quad (2)$$

where λ_d is the known tracer decay constant. For tracers in which k_4 is small compared to the other rate parameters, the tracers become effectively trapped and the process can be assumed to be irreversible. With the assumption that only one homogeneous tissue region is measured and no blood fraction is present, the observed PET data timecourse $TAC(t)$ is the summation of the two compartments $C_f(t)$ and $C_b(t)$ for the tissue:

$$TAC(t) = C_f(t) + C_b(t). \quad (3)$$

Assuming irreversible tracers (i.e., $k_4 = 0$) and solving the differential equations (1,2), we find that:

$$TAC(t) = O(t, \phi) * C_p(t), \quad (4)$$

where $\phi = [k_1 \ k_2 \ k_3]$, $*$ denotes convolution, and

$$O(t, \varphi) = \frac{k_1}{k_2 + k_3} \left(k_3 e^{-\lambda_d t} + k_2 e^{-(k_2 + k_3 + \lambda_d)t} \right) \quad (5)$$

is the system response.

B. Mixture Tissue Model

The previous section described the kinetic model for a single homogeneous tissue component. We now consider a mixture model for heterogeneous tissue that includes Q different components including tracer in the blood. Assuming that we know the plasma input function $C_p(t)$, the TAC for the q^{th} tissue can be written as

$$TAC^q(t) = \frac{k_1^q}{k_2^q + k_3^q} \left(k_3^q e^{-\lambda_d t} + k_2^q e^{-(k_2^q + k_3^q + \lambda_d)t} \right) * C_p(t). \quad (6)$$

For each voxel, the measured TAC is expressed as the linear combination of these tissue TACs:

$$TAC_j(t) = \sum_{q=1}^{Q-1} TAC^q(t) A_{qj} + C_p(t) A_{Qj}, \quad (7)$$

where the voxel index j ranges from 1 to N , and N is the number of voxels. The linear mixing coefficients A_{qj} represent the fractions of different types of metabolically active tissues present in the voxel. Since they are fractions, the A_{qj} should satisfy the constraints that

$$A_{qj} \geq 0, \quad \sum_{q=1}^Q A_{qj} \leq 1. \quad (8)$$

Note that A_{qj} need not strictly sum to 1, since not all voxels will be fully populated by the dynamic components of our model.

III. Model stability

While the mixture model (7) has been used before (e.g., [21], [25]), here we explore the use of additional constraints that are needed to enable accurate parameter estimates because of the low SNR of PET and the ill-conditioning of the inverse problem. Note that our proposed mixture model is closely related to the exponential mixture model, which is well-known to be severely ill-posed [29].

In this section, we consider the estimation of the parameters of (7) from an estimation-theoretic perspective. Our results suggest that the use of signal sparsity information can be used to overcome the ill-posedness of the estimation problem, which motivates the new method we propose in section IV.

A. Cramér-Rao Lower Bounds

The mathematical tool we use to study the model stability is the Cramér-Rao lower bound (CRLB)[30]. The CRLB is defined as the inverse of the Fisher information matrix (FIM):

$$CRLB(\boldsymbol{\theta}) = FIM(\boldsymbol{\theta})^{-1}, \quad (9)$$

where $FIM(\boldsymbol{\theta}) = -E \left[\frac{\partial^2 L(\mathbf{x}|\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right]$, and $L(\mathbf{x}|\boldsymbol{\theta})$ is the log-likelihood function that depends on the measurement vector $\mathbf{x}$ and the parameter vector $\boldsymbol{\theta}$. A well-known result is that for any unbiased estimator $\hat{\boldsymbol{\theta}}$ of $\boldsymbol{\theta}$, the covariance of $\hat{\boldsymbol{\theta}}$ is lower-bounded by the CRLB (with respect to the Loewner ordering of symmetric positive semidefinite matrices):

$$E \left[\left(\hat{\boldsymbol{\theta}} - \mathbf{E}[\hat{\boldsymbol{\theta}}] \right) \left(\hat{\boldsymbol{\theta}} - \mathbf{E}[\hat{\boldsymbol{\theta}}] \right)^T \right] \geq CRLB(\boldsymbol{\theta}). \quad (10)$$

For our stability analysis, we assume that the measured data are modeled as

$$Y_j(t_i) = TAC_j(t_i) + g_{ij} \quad (11)$$

for $i = 1, 2, \dots, I$ and $j = 1, 2, \dots, N$, where I is the number of temporal samples (spread uniformly in time), and g_{ij} is additive white Gaussian noise with variance σ^2 . To simplify notation, we group the model parameters by defining $\boldsymbol{\theta} = [\sigma, \mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_N, \boldsymbol{\phi}_1, \boldsymbol{\phi}_2, \dots, \boldsymbol{\phi}_{Q-1}]$ where $\mathbf{a}_j = [A_{1j}, A_{2j}, \dots, A_{Qj}]$ and $\boldsymbol{\phi}_q = [k_1^q, k_2^q, k_3^q]$.

$$P(\mathbf{Y}|\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{IN}{2}} \sigma^{\frac{IN}{2}}} e^{-\frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^N g_{ij}^2}. \quad (12)$$

Consequently the log-likelihood function (dropping irrelevant additive constants) can be written as:

$$L(\mathbf{Y}|\boldsymbol{\theta}) = -\frac{IN}{2} \log \sigma - \frac{1}{2\sigma^2} \sum_{i=1}^I \sum_{j=1}^N g_{ij}^2. \quad (13)$$

Similar to the derivation in [31], the FIM can then be written as a block matrix as follows:

$$J = \frac{1}{\sigma} \begin{pmatrix} \frac{IN}{2\sigma} & 0 & 0 \\ 0 & \mathbf{\Lambda} & \mathbf{\Delta} \\ 0 & \mathbf{\Delta}^T & \mathbf{\Gamma} \end{pmatrix}, \quad (14)$$

where $\mathbf{\Lambda}$ is a block diagonal matrix with elements

$$\mathbf{\Lambda}_{(j-1)(Q-1)+q, (j'-1)(Q-1)+q'} = \delta(j-j') \sum_{i=1}^I \frac{\partial g_{ij}}{\partial A_{qj}} \frac{\partial g_{ij}}{\partial A_{q'j'}}, \quad (15)$$

$\delta(j)$ is the Kronecker delta function,

$$\Delta_{(j-1)(Q-1)+q, 3*(q-1)+l} = \sum_{i=1}^I \frac{\partial g_{ij}}{\partial A_{qj}} \frac{\partial g_{ij}}{\partial k_l^q}, \quad (16)$$

and

$$\Gamma_{3*(q-1)+l, 3*(q'-1)+l'} = \sum_{i=1}^I \sum_{j=1}^N \frac{\partial g_{ij}}{\partial k_l^q} \frac{\partial g_{ij}}{\partial k_{l'}^{q'}}. \quad (17)$$

The CRLB for the different categories of parameter can then be computed based on the matrix inversion lemma:

$$CRLB(\sigma) = \frac{2\sigma^2}{MN} \quad (18)$$

$$CRLB(\mathbf{A}) = \sigma \left[\mathbf{A} - \mathbf{\Delta} \mathbf{\Gamma}^{-1} \mathbf{\Delta}^T \right]^{-1} \quad (19)$$

$$CRLB(\varphi) = \sigma \left[\mathbf{\Gamma} - \mathbf{\Delta}^T \mathbf{A}^{-1} \mathbf{\Delta} \right]^{-1}. \quad (20)$$

B. Transformed parameterizations

While the choice of $\phi = [k_1, k_2, k_3]$ is natural based on the compartmental model, transformations of these parameters are more relevant for assessing the stability of the estimation problem. In particular, we introduce the new variables $c = \frac{k_1 k_3}{k_2 + k_3}$, $d = \frac{k_1 k_2}{k_2 + k_3}$, and $\theta = k_2 + k_3 + \lambda_d$ and let $\phi = [c, d, \theta]$. With these choices, the response $O(t, \phi)$ in (5) becomes:

$$O(t, \phi) = ce^{-\lambda_d t} + de^{-\theta t} \quad (21)$$

This transform is an invertible transform so that k_1, k_2, k_3 are uniquely determined by c, d and θ .

$$\begin{aligned} k_1 &= c + d \\ k_2 &= \frac{c}{c+d} (\theta - \lambda_d) \\ k_3 &= \frac{d}{c+d} (\theta - \lambda_d) \end{aligned} \quad (22)$$

In this work, we study the stability of the model with respect to the transformed parameterization $\phi = [c, d, \theta]$.

There are various factors that influence the stability of parameter estimation. These include the temporal sampling pattern of the measurement, noise in the blood input function, the number of tissue types Q , the kinetic rate parameters for each mixture component, the number of voxels, the distribution of the fractions A_{qj} and the prior information used in the reconstruction. Combined CRLB-based analysis of all of these factors is beyond the scope of this paper, and the following subsections will focus on the use of prior information

relevant to our proposed approach. For simplicity we assume that we have perfect prior knowledge of the blood input function.

C. Effect of the tissue fraction distribution

In this section, we study how the distribution of the mixing coefficients A_{qj} can influence the stability of parameter estimation. For simplicity, we use only two types of tissues ($Q=3$) with the blood fraction $A_{Qj} = 0$ for all voxels.

In addition, we assume the blood input function satisfies the following parametric model [32]:

$$C_p(t) = (at - b) e^{-\lambda_{b1}t} + be^{-\lambda_{b2}t}. \quad (23)$$

We let $a = 0.0036$ ($\mu\text{Ci/ml}$), $b = 0.0082$ ($\mu\text{Ci/ml}$), $\lambda_{b1} = 0.067$ (Hz), $\lambda_{b2} = 0.0003$ (Hz), which are values we estimated from the blood input function from a set of real patient data. The blood input function is shown in Fig. 2.

The stability of parameter estimation could be significantly influenced by how the fraction parameters A_{qj} are distributed. If the A_{qj} are the same for each voxel, then having access to multiple voxels is equivalent to averaging. On the other hand, if the A_{qj} are distinct for each voxel, then the combination of all voxels could have more or less information about the different mixture parameters than could be obtained by averaging a single voxel N times. In the extreme case where the A_{qj} are binary, then a mixture model is not necessary to represent the signal, and the parameters could be estimated more stably using a single-component model for each voxel. To study these issues, we generated synthetic data where the volume fractions in each voxel for the first mixture component A_{1j} were generated according to either the symmetric Bernoulli distribution or the symmetric Dirichlet distribution [33]. The volume fractions for the second component were obtained from $A_{2j} = 1 - A_{1j}$. The probability density functions for the symmetric Bernoulli distribution and the symmetric Dirichlet distribution are, respectively:

$$f(x) = \frac{1}{2} \delta(x) + \frac{1}{2} \delta(1-x) \quad (24)$$

$$f(x) = \frac{x^{\alpha-1}(1-x)^{\alpha-1}}{\text{Beta}(\alpha, \alpha)}. \quad (25)$$

The Bernoulli distribution leads to binary volume fractions, while the symmetric Dirichlet distribution has a parameter $\alpha \in [0, 1]$ that determines its characteristics. The symmetric Dirichlet distribution is equivalent to the uniform distribution when $\alpha = 1$, but yields a “sparse” distribution (with strong peaks in the PDF near 0 and 1) as α approaches 0. Plots of the Dirichlet distribution are shown for different values of α in Fig. 3

In Fig. 4, we plot CRLBs for θ_1 (the θ parameter for the first mixture component) for both of these different fraction distribution models. The CRLBs were computed holding the kinetic parameters c , d , and θ_1 fixed, while varying θ_2 . These CRLBs are computed assuming no

prior knowledge of the fraction distributions. Note when $\theta_1 = \theta_2$ the two tissues are not distinguishable from each other and the CRLB becomes infinite, as expected based on our previous assessment of this model [31].

The results in Fig. 4 show that the CRLB improves by a small amount as the fraction distributions more heavily favor sparsity. For reference, we also computed CRLBs for the single tissue model for a case where we have pure tissues (i.e., A_{1j} is either 0 or 1) and we have perfect knowledge of the fractions for each voxel. This CRLB is plotted as the black dot-dashed curve in Fig. 4. We observe that the mixture model has a substantially higher CRLB than the single tissue model, even in the case when the mixture fractions are either 0 or 1 (i.e., the Bernoulli case). This is especially true for the region in which the two tissue types have similar rate parameters (i.e., $\theta_1/\theta_2 \approx 1$).

D. Effect of prior knowledge of the tissue fraction distribution

The previous section demonstrated the effects of the tissue fraction distribution for a case in which prior information about the PDF of the tissue fraction distribution was unavailable to the estimator. In this section we study how this prior information can help improve estimation of the mixture model. Instead of estimating both the fractions and the rate parameters, we assume we know the PDF used to generate the fractions and estimate the rate parameters after removing the fractions from the likelihood through marginalization. Given fractions, the likelihood function is:

$$P(Y_j(t_i) | \sigma^2, A_{1j}=x, A_{2j}=1-x, \theta_1, \theta_2) = \frac{1}{(2\pi)^{\frac{1}{2}} \sigma} e^{-\frac{1}{2\sigma^2} (TAC_j(t_i) - TAC^1(t_i, \theta_1)x - TAC^2(t_i, \theta_2)(1-x))^2}, \quad (26)$$

where σ^2 is the noise variance.

Assuming the prior fraction distribution $P(A_{1j}=x, A_{2j}=1-x) = f(x)$ and marginalizing the fraction parameters, we obtain the following expression for the marginalized likelihood function:

$$P(Y_j(t_i) | \sigma^2, \theta_1, \theta_2) = \int_0^1 f(x) P(TAC_j(t_i) | \sigma^2, A_{1j}=x, A_{2j}=1-x, \theta_1, \theta_2) dx. \quad (27)$$

We calculate the CRLBs for the case where the tissue fraction distribution is known using (27). The results are shown in the dashed curves of Fig. 5. From Fig. 5 we can see that the use of prior information can significantly improve the CRLBs compared to estimation without prior information. In addition, it is observed that the performance gain is more significant when the rate constants are similar to each other (i.e. the ratio θ_1/θ_2 is around 1).

We can gain a more intuitive understanding of the importance of using prior information for the mixture model by plotting CRLBs as a function of N . As can be seen in Fig. 6, the mixture model requires roughly 50 $\times$ larger N to achieve the same CRLB as the single tissue model. With prior information, the mixture model only requires roughly 5 $\times$ larger N .

IV. Mixture model parameter Estimation

The previous section demonstrated that the use of constraints is important for the stable estimation of mixture model parameters. In this section, we describe a new constrained approach to PET mixture model estimation that uses a combination of sparsity constraints, spatial smoothness constraints, and a constraint on number of tissue types Q . We keep the transformed parameterization described in section III-B to simplify the description and the optimization. Combining Eqs. (4), (7), and (21), we have

$$TAC_j(t) = \left(\sum_{q=1}^{Q-1} c_q A_{qj} \right) e^{-\lambda_d t} * C_p(t) + \sum_{q=1}^{Q-1} d_q A_{qj} e^{-\theta_q t} * C_p(t) + A_{Qj} C_p(t). \quad (28)$$

In equation (28), Q , c_q , d_q , A_{qj} , and θ_q are unknowns to be estimated, under the constraints that

$$\theta_q > \lambda_d, \quad c_q > 0, \quad d_q > 0, \quad A_{qj} \geq 0, \quad \sum_q A_{qj} \leq 1, \quad (29)$$

and that A_{qj} should be spatially smooth for each q .

It should be noted that the solution to (28) is non-unique, as a result of the presence of non-active tissue compartments, in addition to the bilinearity of A_{qj} with both c_q and d_q . In particular, the signal model is unchanged if A_{qj} is scaled down by a multiplicative factor and the corresponding parameters c_q and d_q are scaled up by the same factor. This

nonuniqueness is a result of the fact that we are allowing $\sum_q A_{qj}$ to be less than one, with $1 - \sum_q A_{qj}$ representing the fraction of non-active tissue compartments within a voxel. To resolve this ambiguity, we impose a constraint on the non-active tissue compartments. In particular, we make the assumption that non-active tissue compartments are not present in all voxels and penalize voxels with non-zero values of $1 - \sum_q A_{qj}$.

The joint estimation of all these parameters is complicated because of nonlinearity, the fact that the model complexity changes significantly for different values of Q , the bilinearity of A_{qj} with both c_q and d_q , and the constraints. As a result, we propose a two stage method to separate the estimation of Q and the θ_q from the remaining variables. In the first stage, we use a sparsity-constrained dictionary approach to simultaneously estimate Q and θ_q . In the second stage we use a block coordinate descent (BCD) algorithm together with a manifold search to robustly estimate the tissue fractions A_{qj} and the linear tissue TAC parameters c_q and d_q .

A. Stage 1: Dictionary-Based Estimation of Q and θ_q

By a change of variables, (28) can be rewritten as

$$TAC_j(t) = B_{0j} e^{-\lambda_d t} * C_p(t) + \sum_{q=1}^{Q-1} B_{qj} e^{-\theta_q t} * C_p(t) + B_{Qj} C_p(t) \quad (30)$$

where we have used $B_{0j} = \sum_{q=1}^{Q-1} c_q A_{qj}$ and $B_{qj} = d_q A_{qj}$ for $q = 1, \dots, Q-1$, and $B_{Qj} = A_{Qj}$.

This change of variables complicates the form of the constraints in (29). However, while these constraints will play important roles in estimating c_q , d_q , and A_{qj} , they are not as essential for estimating Q and θ_q . As a result, we neglect these constraints on B_{qj} for this first stage to ensure that the estimation problem is tractable. Furthermore, equation (30) is a nonlinear model with respect to θ_q and the model complexity is dependent on Q . These facts mean that the direct estimation of θ_q and Q is a difficult problem. To transform equation (30) to a simpler linear model, we use a discretized dictionary-based approach inspired by [22], [23]. In particular, we express (28) using the overcomplete series representation

$$TAC_j(t) = \sum_{n=0}^{N_b} B_{nj} \psi_n(t), \quad (31)$$

where $N_b \gg Q$, and with

$$\psi_0(t) \stackrel{\text{def}}{=} C_p(t), \quad \psi_1(t) \stackrel{\text{def}}{=} e^{-\lambda_d t} * C_p(t), \quad \text{and} \quad \psi_n(t) \stackrel{\text{def}}{=} e^{-\theta_n t} * C_p(t), \quad 2 \leq n \leq N_b. \quad (32)$$

The set of $N_b - 1$ different values of θ_n are selected logarithmically from a physiologically meaningful range such that $\lambda_d < \theta_n < \lambda_{b1}$, where we assume that the rate parameter in any of the tissue types is faster than the exponential decay of the tracer and slower than the fastest rate parameter in our blood input function model (23).

In practice, dynamic PET time series are often reconstructed at a finite number I of time frames, with each frame corresponding to the average tracer behavior over a given time interval. To model this process, we introduce new quantities

$$Y_{ij} = \frac{\int_{t_i^s}^{t_i^e} TAC_j(t) dt}{t_i^e - t_i^s}, \quad \Psi_{in} = \frac{\int_{t_i^s}^{t_i^e} \psi_n(t) dt}{t_i^e - t_i^s} \quad (33)$$

where t_i^s and t_i^e are the sequences of start and end frame times, after which the measured data at the i^{th} time frame in the j^{th} voxel can be modeled as

$$Y_{ij} \approx \sum_{n=0}^{N_b} \Psi_{in} B_{nj} \quad (34)$$

for each i . Equation (34) can equivalently be expressed in matrix form for all voxels and all time points as

$$\mathbf{Y} = \mathbf{\Psi} \mathbf{B} \quad (35)$$

with $\mathbf{Y} \in \mathbb{R}^{I \times N}$, $\mathbf{\Psi} \in \mathbb{R}^{I \times (N_b+1)}$, and $\mathbf{B} \in \mathbb{R}^{(N_b+1) \times N}$.

To solve for $\mathbf{B}$ in equation (35), we define the regularized cost function:

$$cost(\mathbf{B}) = \|\mathbf{Y} - \Psi\mathbf{B}\|_{Fro}^2 + \rho R_{TS}(\mathbf{B}_{2:N_b}, :) + \kappa R_{JS}(\mathbf{B}_{2:N_b}, :) \quad (36)$$

where $\mathbf{B}_{2:N_b}, :$ represents the submatrix of $\mathbf{B}$ with the first two rows removed. We choose

$$R_{TS}(\mathbf{B}) = \sum_{j=1}^N \|\mathbf{B}_{:,j}\|_{\ell_0} \quad (37)$$

where $\mathbf{B}_{:,j}$ represents the j^{th} column of the matrix $\mathbf{B}$, and

$$R_{JS}(\mathbf{B}) = \|\mathbf{B}\|_{row-\ell_0}, \quad (38)$$

where the $row-\ell_0$ quasi-norm of a matrix is defined as the number of non-zero rows [34]. The purpose of $R_{TS}(\mathbf{B})$ is to impose prior information that the tissue fractions in each voxel are frequently sparse, which we saw from section III can substantially improve estimation quality. The purpose of using $R_{JS}(\mathbf{B})$ is that, based on our original modeling assumptions (28), we expect that B_{nj} in (30) will have joint-sparse characteristics, i.e., that the number of n values for which *any* $B_{nj} = 0$ will be very small (and equal to $Q + 1$). The most direct measure of this kind of joint sparsity is the $row-\ell_0$ quasi-norm. Since the first two rows of $\mathbf{B}$ are always present (corresponding to $\psi_0(t)$ and $\psi_1(t)$), we impose the sparsity constraints only on a submatrix of $\mathbf{B}$ with the first two rows removed.

Optimization of equation (36) is NP-hard, but can still be solved in a globally optimal manner if we restrict Q to be small and use a combinatorial algorithm. In particular, assume we are given Q_{max} as the maximum possible value of Q and the matrix Ψ with columns ψ_i . We solve equation (36) by exhaustively searching over all possible nonzero column combinations of Ψ . In particular, we estimate Q and the corresponding columns of Ψ with non-zero coefficients using the following algorithm:

```

Initialize  cost = ∞;
for  Q' = 1:Qmax  do
for  Ψ' ∈ ΩQ'(Ψ)  do
  B' = argminB' ||Y - Ψ'B'||Fro2 + κQ' + ρRTS(B'2:Q', :)
  if  (||Y - Ψ'B'||Fro2 + κQ' + ρRTS(B'2:Q', :) < cost)  (39)
  then
    cost = ||Y - Ψ'B'||Fro2 + κQ' + ρRTS(B'2:Q', :);
    Ψ' = Ψ';
    Q̂ = Q';

```

In this algorithm, we have used the symbol $\Omega_Q(\Psi)$ to denote the set of matrices obtained by combining the first two columns of Ψ with an additional $Q - 1$ columns Ψ ($\Omega_Q(\Psi) = \{[\psi_0, \psi_1, \psi_{i_1}, \psi_{i_2}, \dots, \psi_{i_{Q-1}}]\}$ for all possible monotonically increasing i_1 through i_{Q-1}) and $\mathbf{B}'$ is a $Q + 1$ by N matrix. The θ values used to generate the columns of $\hat{\Psi}'$ are the resulting optimal

estimates of the nonlinear tissue parameters $\hat{\theta}_q$. Equation (39) was solved using the following procedure for each voxel independently:

```

for  $j=1:N$  do
  Initialize  $cost=\infty$ ;
  for  $Q'=1:Q$  do
    for  $\Psi'' \in \Omega_{Q'}(\Psi)$  do
       $\hat{\mathbf{b}}' = \underset{\mathbf{b}}{\operatorname{argmin}} \|\mathbf{Y}_{:,j} - \Psi'' \mathbf{b}'\|_{Fro}^2$  using non-negative least squares (NNLS) (35)
      if  $\|\mathbf{Y}_{:,j} - \Psi'' \hat{\mathbf{b}}'\|_{Fro}^2 + \rho Q' < cost$  then
         $cost = \|\mathbf{Y}_{:,j} - \Psi'' \hat{\mathbf{b}}'\|_{Fro}^2 + \rho Q'$ ;
       $\hat{\mathbf{B}}'_{:,j} = \hat{\mathbf{b}}'$ ;

```

Because we use a combinatorial algorithm, the algorithm parameter κ and ρ have almost no impact on the speed of the algorithm. The computation time of the first stage of the algorithm is determined by Q_{max} and N_b instead.

1) Convex versus Nonconvex Sparsity Penalties—In the sparse reconstruction literature, nonconvex penalty functions are often replaced with convex relaxations such as:

$$R_{JS}(\mathbf{B}) = \|\mathbf{B}\|_{s,t} \stackrel{def}{=} \sum_{\omega}^Q \left[\sum_k^N |\mathbf{B}_{\omega k}|^t \right]^{s/t}, \quad (40)$$

Equation (40) is known to be convex when $s \geq 1$ and $t \geq 1$ while it is also known to encourage row-sparse matrices $\mathbf{B}$ if $s \leq 1$ [36]. As a result, equation (40) seems like a potentially useful surrogate for equation (38) that would enable more computationally-efficient algorithms.

In our experience, convex relaxation methods are significantly outperformed by methods that solve the original nonconvex problem in this application context, due to the significant temporal coherence between the different bases ψ_q . In Fig. 7, we show a simulation that demonstrates the relative performance of the convex penalty in (40) versus our proposed nonconvex penalty function in (38). The results show that the penalty in (40) does lead to row-sparse solutions, with the degree of sparsity controlled by varying the hyperparameter κ . But even with low noise data, the estimated row support does not match well with the true row support for any κ . However, we obtain much better results if we use combinatorial methods for optimization and solve the problem with the nonconvex penalty from (38). We observe from the results in Fig. 7 and related experiments that if N is large enough and if the SNR is relatively high, then we frequently obtain perfect estimates of the row support by using *row- ℓ_0* regularization.

2) Validation that Sparsity Priors Improve Performance—To validate our observations from the CRLB analysis and to test the effectiveness of the tissue sparsity regularization term $R_{TS}(\mathbf{B})$, we also performed a simulation in which we optimized (36) with

different values of the hyperparameter ρ . The results are compared based on the root mean square error percentage (RMSEP) of the estimated rate parameters, with

$$RMSEP(\hat{\theta}_1, \hat{\theta}_2) = \frac{\sqrt{(\hat{\theta}_1 - \theta_1)^2 + (\hat{\theta}_2 - \theta_2)^2}}{\sqrt{\theta_1^2 + \theta_2^2}}. \quad (41)$$

Results, Fig. 8, show that the use of sparsity regularization improves the RMSEP. Additional simulations (not shown) demonstrate that the sparse tissue penalty term $R_{TS}(\mathbf{B})$ becomes more helpful as the true fraction distribution becomes closer to a binary distribution.

B. Stage 2: Estimation of $\mathbf{c}_q$, $\mathbf{d}_q$ and $\mathbf{A}_{qj}$

Solving (36) provides us with estimated $\hat{Q}$ and $\hat{\theta}_q$ values, and with these parameters fixed, we return to solving (28) for the remaining parameters. Defining

$$\mathbf{c} = [c_1, c_2, \dots, c_{\hat{Q}-1}] \quad \text{and} \quad \mathbf{D} = \text{diag}\{d_1, d_2, \dots, d_{\hat{Q}-1}\}, \quad (42)$$

our model can be expressed in matrix form as

$$\begin{aligned} \mathbf{Y} &\approx \hat{\Psi}' \begin{pmatrix} \mathbf{c} & 0 \\ F & \mathbf{0} & 1 \\ \mathbf{D} & \mathbf{0} \end{pmatrix} \mathbf{A} \\ &= \hat{\Psi}' \mathbf{F} \mathbf{A}, \end{aligned} \quad (43)$$

where $\mathbf{A} \in \mathbb{R}^{Q \times N}$ is the matrix of A_{qj} , and $\mathbf{F} \in \mathbb{R}^{(\hat{Q}+1) \times \hat{Q}}$ is given by

$$\mathbf{F} = \begin{pmatrix} \mathbf{c} & 0 \\ \mathbf{0} & 1 \\ \mathbf{D} & \mathbf{0} \end{pmatrix}. \quad (44)$$

There are $2(\hat{Q} - 1)$ unknown parameters (c_q, d_q) in $\mathbf{F}$ and $N\hat{Q}$ unknown parameters in $\mathbf{A}$. In addition, even with use of the constraints in (29), the least squares fit of these parameters would not be unique because of bilinearity between $\mathbf{F}$ and $\mathbf{A}$, which comes from the fact that $\alpha\mathbf{F}$ and $\frac{1}{\alpha}\mathbf{A}$ have the same residual as $\mathbf{F}$ and $\mathbf{A}$, where α is any positive number. We need additional prior information to uniquely and stably solve for $\mathbf{F}$ and $\hat{\mathbf{A}}$. In addition to (29), we also apply:

1. Sparsity-enforcing ℓ_1 regularization on $\left(1 - \sum_q A_{qj}\right)$ to impose the prior information that not all of the voxels have metabolically inactive tissue present. This constraint removes the bilinear scale ambiguity between $\mathbf{F}$ and $\mathbf{A}$.
2. Quadratic smoothness regularization of the tissue fractions A_{qj} , to impose the prior information that tissue contributions will often vary smoothly in standard PET images.

Combining these constraints, we define our cost function for the second stage as:

$$cost(\mathbf{F}, \mathbf{A}) = \|\mathbf{Y} - \hat{\Psi}' \mathbf{F} \mathbf{A}\|_{Fro}^2 + \beta \sum_q \sum_j^N \sum_{l \in Nbr(j)} (A_{qj} - A_{gl})^2 + \gamma \sum_j^N |1 - \sum_{q=1}^{\hat{Q}} A_{qj}|, \quad (45)$$

where $\mathbf{F} \in \Omega_F = \left\{ \mathbf{F} : \mathbf{F} = \begin{pmatrix} \mathbf{c} & \mathbf{0} \\ \mathbf{0} & \mathbf{1} \\ \mathbf{D} & \mathbf{0} \end{pmatrix}, \mathbf{D} = diag(\mathbf{d}), c_q > 0, d_q > 0 \right\}$, and

$\mathbf{A} \in \Omega_A = \left\{ \mathbf{A} : A_{gj} \geq 0, \sum_q A_{gj} \leq 1 \right\}$. $Nbr(j)$ denotes the set of neighboring voxels for voxel j , and γ and β are regularization parameters.

The parameter γ removes the bilinear scale ambiguity between $\mathbf{F}$ and $\mathbf{A}$. We have found that for practical applications, the results are relatively insensitive to changes in γ over large range (γ values between 10^{-8} and 10^{-2} all produced nearly identical results to the reconstructions shown later in the paper). Outside this range high bias can occur depending on the amount of zero activity tissue in the FOV. The parameter β is similar to the spatial smoothness regularization used in [37] and [38], and can be tuned by the user to achieve a reasonable balance between noise and spatial resolution of the estimated fraction images.

To find the optimal solution to equation (45), we alternate between estimating $\mathbf{F}$ and $\mathbf{A}$ using a block coordinate descent (BCD) approach (i.e., alternating between optimization over $\mathbf{F}$ with $\mathbf{A}$ fixed, and optimization over $\mathbf{A}$ with $\mathbf{F}$ fixed). However, although our regularization framework eliminates potential bilinear scale ambiguity between the optimal values of $\mathbf{F}$ and $\mathbf{A}$, the bilinear nature of the problem still reduces the convergence rate of our alternating algorithm when the current estimates of $\mathbf{F}$ and $\mathbf{A}$ lie in a region of the parameter space where

$\mathbf{Y} \approx \hat{\Psi}' \mathbf{F} \mathbf{A}$ but $(1 - \sum_q A_{gj})$ is not optimally sparse. This is because the alternating BCD method does not allow for simultaneous update of $\mathbf{A}$ and $\mathbf{F}$, leading to slow convergence if $\mathbf{F}$ and $\mathbf{A}$ need to be concurrently rescaled. To overcome this problem, we introduce an additional manifold search into our BCD algorithm to resolve problems of scale ambiguity in this region of the parameter space. A sketch of the full algorithm is given below:

1. Initialize $\mathbf{A}$ with uniform images ($A_{gj} = 1/Q$).
2. At the m^{th} iteration, set $\mathbf{A}$ equal to the estimate from the previous iteration $\hat{\mathbf{A}}_{m-1}$ and solve:

$$\hat{\mathbf{F}}_m = \underset{\mathbf{F} \in \Omega_F}{argmin} \quad cost(\mathbf{F}, \hat{\mathbf{A}}_{m-1}) = \|\mathbf{Y} - \hat{\Psi}' \mathbf{F} \hat{\mathbf{A}}_{m-1}\|_{Fro}^2 \quad (46)$$

using NNLS [35].

3. For fixed $\hat{\mathbf{F}}_m$, solve

$$\hat{\mathbf{A}} = \underset{\mathbf{A} \in \Omega_A}{\operatorname{argmin}} \operatorname{cost}(\hat{\mathbf{F}}_m, \mathbf{A}) = \underset{\mathbf{A} \in \Omega_A}{\operatorname{argmin}} \left\| \left(\mathbf{Y} - \hat{\Psi}' \hat{\mathbf{F}}_m \mathbf{A} \right) \right\|_{Fro}^2 + \beta \sum_{q=1}^{\hat{Q}} \sum_{l=1}^N \sum_{l \in N_{Br}(j)} (A_{ql} - A_{ql})^2 + \gamma \sum_j^N \left| 1 - \sum_{q=1}^{\hat{Q}} A_{qj} \right| \quad (47)$$

using an Alternating Direction Method of Multipliers (ADMM) algorithm [39]. The details of our ADMM implementation are given in the appendix.

4. To resolve potential scale ambiguity between vector $\boldsymbol{\eta} \in \mathbb{R}^{\hat{Q}-1}$ and perform a manifold search. Defining $\Lambda \boldsymbol{\eta} = \operatorname{diag}\{1, \boldsymbol{\eta}\}$ we note that

$\hat{\mathbf{F}}_m \hat{\mathbf{A}}_m = (\hat{\mathbf{F}}_m \Lambda_{\boldsymbol{\eta}}^{-1}) (\Lambda_{\boldsymbol{\eta}} \hat{\mathbf{A}}_m)$ for any strictly positive choices for the elements of $\boldsymbol{\eta}$. This implies that by modifying the values of $\boldsymbol{\eta}$, we can shift the scaling between $\hat{\mathbf{F}}_m$ and $\hat{\mathbf{A}}_m$ without modifying the data consistency term of (45). Thus, we calculate

$$\boldsymbol{\eta}_m = \underset{\boldsymbol{\eta} \in \Omega_{\boldsymbol{\eta}}}{\operatorname{argmin}} \operatorname{cost}(\hat{\mathbf{F}}_m \Lambda_{\boldsymbol{\eta}}^{-1}, \Lambda_{\boldsymbol{\eta}} \hat{\mathbf{A}}_m) \quad (48)$$

with $\Omega_{\boldsymbol{\eta}} = \left\{ \boldsymbol{\eta} : \eta_q \geq 0; \eta_q \hat{\mathbf{A}}_{m,gj} \leq 1, \forall j; 1 - \sum_q^{\hat{Q}-1} \eta_q \hat{\mathbf{A}}_{m,qj} - \hat{\mathbf{A}}_{m,\hat{Q}j} \geq 0, \forall j \right\}$.

This is a small scale quadratic programming problem and is solved using a standard interior point method.

5. Set $\hat{\mathbf{F}}_m = \hat{\mathbf{F}}_m \Lambda_{\boldsymbol{\eta}}^{-1}$ and $\hat{\mathbf{A}}_m = \Lambda_{\boldsymbol{\eta}} \hat{\mathbf{A}}_m$.
6. Iterate steps (2)-(5) until convergence.

V. RESULTS

A. Dynamic simulation

We simulated a 2-D scanner with the same geometry as the Siemens mCT scanner with 4 mm detectors. The reconstructed image size was 128×128 with 2 mm voxel size and the sinogram dimension was 400×168 (400 ray samples and 168 angular samples). A total of 30 frames with varying frame duration was simulated for a 60-min dynamic scan (8 frames of 15 seconds, 4 frames of 30 seconds, 11 frames of 1 minute, 5 frames of 5 minutes and 2 frames of 10 minutes). There were a total of 500 million counts across all 30 frames. Each frame was reconstructed using MAP [40] reconstruction with 50 iterations of the algorithm. The spatial smoothness regularization parameter of MAP reconstruction was set to 0.1. The dynamic images were prewhitened for unit variance based on the variance estimated using the method in [37]. Spatial correlation between voxels was ignored.

We simulated two dynamic image sequences, each representing tumors at different stages. Images of the true fractions are shown in the top halves of Figs. 9 and 10. Fig. 9 shows the simulation of 4 small metastases that are at an early stage of development, where the tumor size is small and homogeneous, and necrosis has not yet developed. Fig. 10 shows the simulation of 4 tumors at a late stage of development where the tumors are large and heterogeneous, and active tumor cells have formed donut shapes around necrotic centers. In both simulations, the tumors are embedded in a background of normal tissue, with a large

simulated blood vessel at the center of the ROI. We applied our proposed method with $\kappa = 1 \times 10^{-2}$, $\rho = 1 \times 10^{-4}$, $\beta = 1.2 \times 10^{-5}$, and $\gamma = 1.3 \times 10^{-5}$ to the simulated image sequences. The regularization parameters were chosen as follows. In the first stage, we chose κ such that $Q=3$. We chose ρ so that 20% of the image voxels contained more than one contributing tissue component in the first stage of the algorithm. In the second stage, we chose γ high enough so that zero activity tissue was mostly confined to necrotic cores. We chose β subjectively to obtain a reasonable balance between SNR and resolution. Representative estimated tissue fraction images are shown in the bottom halves of Figs. 9 and 10, and the estimated tissue TACs are shown in Fig. 11 and Fig. 12 with both mean and standard deviation plotted from the results of 50 Monte Carlo simulations. For comparison, we also applied a single tissue model on hand-drawn tumor and normal tissue ROIs (marked in red and green respectively in Figs. 9 and 10).

Our results show that in both cases, by explicitly modelling the tissue heterogeneity, our proposed method provides accurate estimates of the tissue TACs with smaller bias than obtained with the single tissue model, which suffers from low scanner resolution and partial volume effects. The tumor TAC is estimated more accurately for the necrotic tumor simulation than it was in the metastatic tumor simulation. This is because there are many more tumor-containing voxels in the necrotic tumor simulation. For further quantitative evaluation of performance, we estimate tumor volume by summing the estimated tumor volume based on the fractions obtained from the mixture model. For comparison, we also estimate tumor volume based on the estimated tissue parameters from the single-tissue model (using these values as inputs to the second stage of our proposed algorithm). The results of this quantitative comparison are shown in Tables I and II.

Tables I and II show that mixture model yields low bias in the estimated tumor volume while the single tissue model parameter estimates lead to high bias caused by tumor heterogeneity and partial volume effects. While the variance of both methods is comparable, the low bias of the mixture model means that it has a much smaller mean-squared error.

B. Real patient data

A one-hour dynamic FDG PET scan was performed on a patient with colon cancer and liver metastases using a Siemens Biograph PET scanner with 4 mm detectors. Dynamic list mode data were binned into 30 inhomogeneous time frames with increasing frame durations (the frame durations were the same as in the simulation), and reconstructed images were generated using MAP reconstruction with a 2 mm voxel size. The dynamic images were prewhitened for unit variance as in the simulations. Spatial correlation between voxels were ignored. The blood input function was acquired by manually segmenting blood signal in the left ventricle. To limit computation cost, the first stage algorithm of our mixture model was applied to a 2D transaxial slice with $\kappa = 1 \times 10^{-2}$, and $\rho = 1 \times 10^{-4}$. The second stage of our mixture model was applied to the whole 3D volume with $\beta = 3.3 \times 10^{-5}$, and $\gamma = 9.3 \times 10^{-5}$. The parameters were chosen as follows. We chose our hyperparameters in the similar fashion as we did in Section V-A. We chose β subjectively to obtain a reasonable balance between SNR and resolution. Results are shown in Fig. 13 and Fig. 14. Note we only show a

2D transaxial slice (the same slice to which the first stage algorithm was applied) through the 3D volume.

The CT image in Fig. 13(a) shows a number of lesions within the liver (see arrows). From the coregistered static PET images in Fig. 13(d), we see that there is increased activity in these same areas in the liver. In Fig. 14 we show the tissue fraction images estimated using our mixture model, coregistered and overlaid on CT images. From the tumor and normal tissue fraction image, we can see enhanced tumor contrast compared to static PET images. To quantify the difference in contrast, we compute mean values in tumor and normal tissue ROIs (Fig. 13(c)) using both the static PET images and our estimated tumor mixtures. Contrast is defined as the ratio of the mean values in tumor and normal tissue ROIs. The statistics are summarized in Table III, which shows that the proposed method substantially improves the contrast between tumor and normal contrast. Overall the tissue fractions appear consistent with the anatomy and lesions seen in the CT. Note for example the high blood fraction in the abdominal aorta, and the relatively uniform normal tissue regions in the liver where lesions are not visible. The tumor and zero-activity fractions reveal tumors with active boundaries and necrotic (zero-activity) cores.

VI. Discussion

The computation time of the two stage algorithm is mainly limited by the first stage where the nonlinear parameters need to be estimated using combinatorial algorithms. For $Q_{max} = 3$ and $N_b = 50$, we need to evaluate equation (39) for a total of $\binom{50}{1} + \binom{50}{2} = 1275$ times while the second stage of the algorithm typically takes 10-20 iterations to converge. The detailed computation time based on a 256 by 256 2D slice is shown in Table IV. All computations were done on a Windows workstation with one AMD Phenom 9950 Quad processor (2.6GHz) with 16GB of RAM. The algorithm is implemented in MATLAB and is partly multithreaded. Because of the high computational cost for first stage, for 3D data we typically compute the nonlinear tissue parameters based on a 2D slice and then apply the second stage of the algorithm to the whole 3D volume.

In this paper we did not study the effects of an inaccurate blood input function. We assume a parametric model (23) is fitted to sample values derived from the dynamic images. This will lead to some errors in the input function (a combination of bias and variance). We have not yet explored the relative impact of these errors on our approach vs. standard kinetic modeling. Since we jointly estimate two or more sets of kinetic parameters the estimation problem is more illposed (as explored in Section III) so we can expect somewhat increased sensitivity to errors in the input function. Fully exploring the impact of this error is beyond the scope of this paper, but is an interesting topic for future work.

In this paper we did not fully study the impact of hyperparameter β , which enforces spatial smoothness. In our previous work in PET reconstruction [37], we have explicitly linked selection of β to image resolution through the Fisher information matrix. And by using spatial varying β we can achieve approximately uniform resolution. Here in the mixture model, the problem is more difficult since a more complex model is used. Using a fixed β

value across space and tissue fractions, as we do here, will lead to spatially variant resolution. Investigation of the degree of variability and development of methods to produce approximate invariance, analogous to those in [37] and [38], is an interesting topic for future work.

VII. CONCLUSION

In this paper, we proposed and investigated a new approach to estimating PET kinetic parameters with a mixture model. Theoretical analysis using the CRLB and empirical studies demonstrated that the estimation of mixture model parameters is greatly enhanced by incorporating prior knowledge about sparsity. Based on these findings, we proposed a two stage algorithm that successfully estimates tissue kinetic parameters and tissue fractions for simulated and real data with relatively high noise. We anticipate that the estimated tissue fractions could be used to measure tumor heterogeneity, which could in turn be used for tumor staging, treatment assessment and optimization. Our proposed method can be directly generalized to apply in other fields that involve mixture models, such as DCE MRI and DCE CT.

Acknowledgment

This work is supported by NIH grants R21CA149587 and R01-EB010197.

VIII. Appendix

This appendix describes how we solve equation (47) using the ADMM algorithm as in [39]. First, by integrating the constraints into the cost function, we note that constrained optimization problem in equation (47) is equivalent to the unconstrained optimization problem in equation (49).

$$cost_{uc}(\tilde{\mathbf{A}}) = \left\| \left(Y - \tilde{\Psi} \tilde{\mathbf{A}} \right) \right\|_{Fro}^2 + \beta \left\| \tilde{\mathbf{E}} \tilde{\mathbf{A}}' \right\|_{Fro}^2 + \gamma \left| \tilde{\mathbf{A}}_{Q+1,:} \right|_1 + \iota_{\{\tilde{\mathbf{A}}'_{1=1}\}} + \iota_{\{\tilde{\mathbf{A}} \geq 0\}}, \quad (49)$$

where $\iota_{\{H\}}$ is the indicator function of the condition H , that is:

$$\iota_{\{H\}} = \begin{cases} 0, & H \text{ is true} \\ \infty, & H \text{ is false} \end{cases}, \quad (50)$$

$$\tilde{\mathbf{A}} = \begin{pmatrix} \mathbf{A} \\ 1 - \sum_{q=1}^{\hat{Q}} A_q, : \end{pmatrix}, \quad (51)$$

$$\tilde{\Psi} = \begin{pmatrix} \hat{\Psi}' \mathbf{F}_{m-1} & 0 \end{pmatrix}, \quad (52)$$

$$\|\mathbf{E}\mathbf{A}'\|_{Fro}^2 = \sum_j^N \sum_{l \in NBr(j)} (A_{qj} - A_{ql})^2, \quad (53)$$

and

$$\tilde{\mathbf{E}} = (\mathbf{E} \quad \mathbf{0}). \quad (54)$$

If we let:

$$f_1(\tilde{\mathbf{A}}) = \|\mathbf{Y} - \tilde{\Psi}\tilde{\mathbf{A}}\|_{Fro}^2 + \beta \|\tilde{\mathbf{E}}\tilde{\mathbf{A}}'\|_{Fro}^2 + \iota_{\{\tilde{\mathbf{A}}' \mathbf{1} = \mathbf{1}\}} \quad (55)$$

and

$$f_2(\tilde{\mathbf{A}}) = \gamma |\tilde{A}_{Q+1,:}|_1 + \iota_{\{\tilde{\mathbf{A}} \geq \mathbf{0}\}}, \quad (56)$$

then minimizing equation (49) is equivalent to minimizing equation (57) under the constraint that $\mathbf{U} = \tilde{\mathbf{A}}$:

$$\text{cost}(\tilde{\mathbf{A}}, \mathbf{U}) = f_1(\tilde{\mathbf{A}}) + f_2(\mathbf{U}). \quad (57)$$

The ADMM algorithm minimizes the constrained version of (57) by alternating between optimization over $\tilde{\mathbf{A}}$, optimization over $\mathbf{U}$, and updating the Lagrange multipliers for the constraint $\tilde{\mathbf{A}} = \mathbf{U}$. In particular, the update procedure is:

For $k = 1 : K$,

$$\tilde{\mathbf{A}}_{k+1} = \underset{\tilde{\mathbf{A}}}{\text{argmin}} f_1(\tilde{\mathbf{A}}) + \frac{\mu}{2} \|\tilde{\mathbf{A}} - \mathbf{U}_k - \mathbf{\Lambda}_k\|_{Fro}^2 \quad (58)$$

$$\mathbf{U}_{k+1} = \underset{\mathbf{U}}{\text{argmin}} f_2(\mathbf{U}) + \frac{\mu}{2} \|\tilde{\mathbf{A}}_{k+1} - \mathbf{U} - \mathbf{\Lambda}_k\|_{Fro}^2 \quad (59)$$

$$\mathbf{\Lambda}_{k+1} = \mathbf{\Lambda}_k - (\tilde{\mathbf{A}}_{k+1} - \mathbf{U}_{k+1}) \quad (60)$$

end where K is the number of iterations. Although ADMM does not strictly decrease the cost function in equation (49), it converges to the global minimum of this convex subproblem with convergence rate $O(1/k)$.

Equation (58) is a quadratic optimization problem with equality constraints for which we used a projected conjugate gradient method.

Equation (59) also has a closed form solution, given by:

$$U_{q,j} = \begin{cases} \max(\tilde{A}_{q,j} - \Lambda_{q,j}, 0), & q \leq \hat{Q} \\ \max(\text{soft}(\tilde{A}_{q,j} - \Lambda_{q,j}, \gamma/\mu), 0), & q = \hat{Q} + 1 \end{cases} \quad (61)$$

where

$$\text{soft}(a, b) = \begin{cases} a - b, & a \geq b \\ 0, & a < b \end{cases} \quad (62)$$

is a soft thresholding operator [41]. As a result, each step of the algorithm can be computed quickly.

In ADMM, the parameter μ only impacts the convergence rate of the algorithm. A discussion of the theoretical issues underlying selection of this parameter can be found in [42] and [43]. Practically we found that the best parameter for the proposed method was between 10^{-6} and 10^{-4} for both the simulated and in vivo data.

REFERENCES

1. Ritschel, W.; Kearns, G. Handbook of basic pharmacokinetics– including clinical applications. American Pharmacists Association; Washington, DC: 2004.
2. Gunn R, Gunn S, Cunningham V. Positron emission tomography compartmental models. Journal of Cerebral Blood Flow & Metabolism. 2001; 21(6):635–652. [PubMed: 11488533]
3. Ogawa T, Kanno I, Shishido F, Inugami A, Higano S, Fujita H, Murakami M, Uemura K, Yasui N, Mineura K, et al. Clinical value of PET with 18F-fluorodeoxyglucose and L-methyl-11C-methionine for diagnosis of recurrent brain tumor and radiation injury. Acta Radiologica. 1991; 32(3):197–202. [PubMed: 2064862]
4. Dimitrakopoulou-Strauss A, Strauss L, Schwarzbach M, Burger C, Heichel T, Willeke F, Mechttersheimer G, Lehnert T. Dynamic PET 18F-FDG studies in patients with primary and recurrent soft-tissue sarcomas: impact on diagnosis and correlation with grading. Journal of Nuclear Medicine. 2001; 42(5):713–720. [PubMed: 11337565]
5. Lammertsma A, Hume S, et al. Simplified reference tissue model for PET receptor studies. Neuroimage. 1996; 4(3):153–158. [PubMed: 9345505]
6. Slifstein M, Kolachana B, Simpson E, Tabares P, Cheng B, Duvall M, Frankle W, Weinberger D, Laruelle M, Abi-Dargham A. COMT genotype predicts cortical-limbic d1 receptor availability measured with [11C]NNC112 and PET. Molecular psychiatry. 2008; 13(8):821–827. [PubMed: 18317466]
7. Berti V, Polito C, Ramat S, Vanzi E, De Cristofaro M, Pellicano G, Mungai F, Marini P, Formiconi A, Sorbi S, et al. Brain metabolic correlates of dopaminergic degeneration in de novo idiopathic parkinson's disease. European journal of nuclear medicine and molecular imaging. 2010; 37(3):537–544. [PubMed: 19727712]
8. Nye J, Votaw J, Jarkas N, Purselle D, Camp V, Bremner J, Kilts C, Nemeroff C, Goodman M. Compartmental modeling of 11CHOMADAM binding to the serotonin transporter in the healthy human brain. Journal of Nuclear Medicine. 2008; 49(12):2018–2025. [PubMed: 19038997]
9. Yokota J. Tumor progression and metastasis. Carcinogenesis. 2000; 21(3):497. [PubMed: 10688870]
10. Eary J, O'Sullivan F, O'Sullivan J, Conrad E. Spatial heterogeneity in sarcoma 18F-FDG uptake as a predictor of patient outcome. Journal of Nuclear Medicine. 2008; 49(12):1973–1979. [PubMed: 18997052]
11. Kamasak M, Bouman C, Morris E, Sauer K. Direct reconstruction of kinetic parameter images from dynamic PET data. Medical Imaging, IEEE Transactions on. May; 2005 24(5):636–650.

12. Huang S, Zhou Y. Spatially-coordinated regression for image-wise model fitting to dynamic PET data for generating parametric images. *Nuclear Science, IEEE Transactions on*. 1998; 45(3):1194–1199.
13. O'Sullivan F, Saha A. Use of ridge regression for improved estimation of kinetic constants from PET data. *Medical Imaging, IEEE Transactions on*. 1999; 18(2):115–125.
14. Tang J, Kuwabara H, Wong D, Rahmim A. Direct 4d reconstruction of parametric images incorporating anato-functional joint entropy. *Physics in medicine and biology*. 2010; 55:4261. [PubMed: 20647600]
15. Thireou T, Strauss L, Dimitrakopoulou-Strauss A, Kontaxakis G, Pavlopoulos S, Santos A. Performance evaluation of principal component analysis in dynamic fdg-pet studies of recurrent colorectal cancer. *Computerized medical imaging and graphics*. 2003; 27(1):43–51. [PubMed: 12573889]
16. Pedersen F, Bergströme M, Bengtsson E, Långström B. Principal component analysis of dynamic positron emission tomography images. *European Journal of Nuclear Medicine and Molecular Imaging*. 1994; 21(12):1285–1292.
17. Naganawa M, Kimura Y, Ishii K, Oda K, Ishiwata K, Matani A. Extraction of a plasma time-activity curve from dynamic brain pet images based on independent component analysis. *Biomedical Engineering, IEEE Transactions on*. 2005; 52(2):201–210.
18. Lee J, Lee D, Ahn J, Cheon G, Kim S, Yeo J, Seo K, Park K, Chung J, Lee M. Blind separation of cardiac components and extraction of input function from h215o dynamic myocardial pet using independent component analysis. *Journal of Nuclear Medicine*. 2001; 42(6):938–943. [PubMed: 11390560]
19. Ahn J, Kim S, Oh J, Choi S. Multiple nonnegative-matrix factorization of dynamic PET images. *Proceedings of Asian Conference on Computer Vision*. 2004
20. Ahn J, Lee D, Lee J, Kim S, Cheon G, Yeo J, Shin S, Chung J, Lee M. Quantification of regional myocardial blood flow using dynamic h215o pet and factor analysis. *Journal of Nuclear Medicine*. 2001; 42(5):782–787. [PubMed: 11337577]
21. O'Sullivan F. Imaging radiotracer model parameters in PET: a mixture analysis approach. *Medical Imaging, IEEE Transactions on*. Sep; 1993 12(3):399–412.
22. Gunn RN, Gunn SR, Turkheimer FE, Aston JAD, Cunningham VJ. Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling. *J Cereb Blood Flow Metab*. 2002; 22(12):1425–1439. [PubMed: 12468888]
23. Peng J, Aston J, Gunn R, Liou C, Ashburner J. Dynamic positron emission tomography data-driven analysis using sparse bayesian learning. *Medical Imaging, IEEE Transactions on*. 2008; 27(9):1356–1369.
24. Jiang H, Ogden R. Mixture modeling for dynamic PET data. *Statistica Sinica*. 2008; 18(4):1341.
25. O'Sullivan F. Locally constrained mixture representation of dynamic imaging data from PET and MR studies. *Biostatistics*. Apr; 2006 7(2):318–338. [PubMed: 16361274]
26. Lin Y, Li Q, Haldar JP, Leahy RM. Constrained mixture modeling for the estimation of kinetic parameters in dynamic pet. *Biomedical Imaging (ISBI), 2012 9th IEEE International Symposium on*. IEEE. 2012:1004–1007.
27. Godfrey K. *Compartmental models and their application*. 1983
28. Huang S, Phelps M, Hoffman E, Sideris K, Selin C, Kuhl D. Noninvasive determination of local cerebral metabolic rate of glucose in man. *American Journal of Physiology-Endocrinology And Metabolism*. 1980; 238(1):E69–E82.
29. Istratov A, Vyvenko O. *Exponential analysis in physical phenomena. Review of Scientific Instruments*. 1999; 70:1233.
30. Cramér H. *Mathematical methods of statistics*. 1946
31. Lin Y, Li Q, Leahy R. Kinetic parameters estimation for heterogeneous tumor model. *Nuclear Science Symposium Conference Record (NSS/MIC)*. Nov.2009 :3069–3073.
32. Feng D, Huang S, Wang X. Models for computer simulation studies of input functions for tracer kinetic modeling with positron emission tomography. *International journal of bio-medical computing*. 1993; 32(2):95–110. [PubMed: 8449593]

33. Connor RJ, Mosimann JE. Concepts of independence for proportions with a generalization of the dirichlet distribution. *Journal of the American Statistical Association*. 1969; 64(325):194–206.
34. and Tropp JA. Algorithms for simultaneous sparse approximation. part ii: Convex relaxation. *Signal Processing*. 2006; 86(3):589–602.
35. Lawson C, Hanson R. Solving least squares problems. SIAM. 1995; 15
36. Fan J, Li R. Variable selection via nonconcave penalized likelihood and its oracle properties. *Journal of the American Statistical Association*. 2001; 96(456):1348–1360.
37. Qi J, Leahy R. Resolution and noise properties of MAP reconstruction for fully 3-D PET. *Medical Imaging, IEEE Transactions on*. 2000; 19(5):493–506.
38. Fessler JA, Rogers WL. Spatial resolution properties of penalized-likelihood image reconstruction: space-invariant tomographs. *Image Processing, IEEE Transactions on*. 1996; 5(9):1346–1358.
39. Bioucas-Dias J, Figueiredo M. Alternating direction algorithms for constrained sparse regression: Application to hyperspectral unmixing. *Hyperspectral Image and Signal Processing: Evolution in Remote Sensing (WHISPERS), 2010 2nd Workshop on*. Jun.2010 :1–4.
40. Qi J, Leahy R, Cherry S, Chatziioannou A, Farquhar T. High-resolution 3D bayesian image reconstruction using the microPET small-animal scanner. *Physics in Medicine and Biology*. 1998; 43:1001. [PubMed: 9572523]
41. Donoho DL. De-noising by soft-thresholding. *Information Theory, IEEE Transactions on*. 1995; 41(3):613–627.
42. Boley, D. Linear convergence of admm on a model problem. Department of Computer Science and Engineering, University of Minnesota; TR: 2012. p. 12-009.
43. Ghadimi E, Teixeira A, Shames I, Johansson M. Optimal parameter selection for the alternating direction method of multipliers (admm): quadratic problems. *arXiv preprint arXiv:1306.2454*. 2013

Fig. 1.

General three compartment model. $C_p(t)$ is the activity in the blood input function. $C_f(t)$ is the activity of the free tracer in tissue. $C_b(t)$ is the bound tracer.

Fig. 2.
The blood input function used for the CRLB analysis.

Fig. 3. Plots of the symmetric Dirichlet PDF for different values of the parameter α .

Fig. 4.

Relative CRLBs for the rate parameter θ_1 for different fraction distributions and different θ_2 values. All curves were generated assuming $c_1 = 0.0004$ (Hz), $d_1 = 0.003$ (Hz), $c_2 = 0.0005$ (Hz), $d_2 = 0.004$ (Hz), $\theta_1 = 0.0033$ (Hz) and $N = 200$. Each CRLB curve represents the average CRLB from 100 randomly sampled realizations of the fraction distribution. The black dot-dashed curve is the CRLB calculated using a single tissue model when we have pure tissues and perfect knowledge about the fractions for each voxel (100 voxels with $A_{1j} = 0$ and 100 voxels with $A_{1j} = 1$).

Fig. 5.

Dashed curves show CRLBs after introducing prior information about the fraction distribution according to (27). The rest of the curves and parameters are the same as in Fig. 4.

Fig. 6.

CRLB results as a function of the number of voxels N . The black dot-dashed curve represents the CRLB for the single tissue model using $N = 20$. The rest of the curves show CRLBs for different cases of the mixture model, with dashed and solid curves corresponding to cases with and without prior information, respectively. For easy comparison to the single tissue model, each curve is plotted as a function of $N/20$, i.e., the ratio between the number of voxels used in the mixture model versus the number of voxels used to compute the black dot-dashed single tissue CRLB curve.

Fig. 7.

Simulation results showing the estimation of $\mathbf{B}$ using different regularization functions. The estimated $\mathbf{B}$ is shown averaged over all voxels as a function of the rate parameters θ . The simulation used two tissue types with $\theta_1 = 0.0012$ (Hz) and $\theta_2 = 0.023$ (Hz), $N = 4096$, and $SNR = 100$. The cyan curve shows the location of the true rate parameters. The blue, green and red curves show the estimated $\mathbf{B}$ using convex relaxed regularization (40) for different regularization parameters and with $s = t = 1$. The black curve shows the estimated $\mathbf{B}$ using the proposed $row-\ell_0$ regularization (38). The cyan curve and the black curve overlap with each other with slight difference in height.

Fig. 8.

Root mean square error percentage (averaged over 100 Monte Carlo simulation experiments) of the estimated rate parameters as a function of the sparsity-promoting regularization parameter ρ . The simulated data used to generate this plot were all generated using $Q = 3$, with $\theta_1 = 0.0012$ (Hz), $\theta_2 = 0.023$ (Hz), SNR = 100, and $N = 200$. The tissue fractions were sampled from the Dirichlet distribution with parameter $\alpha = 0.1$.

Fig. 9.

Metastatic tumor simulation. The top row corresponds to the true fractions. The bottom row corresponds to the estimated fractions using our proposed method. From left to right, each column corresponds to tumor, normal tissue, blood and necrotic tissue. The red disk shown in (a) marks the boundary of the tumor ROI used for the single tissue model comparison. The green ellipse in (b) marks the boundary of the normal tissue ROI.

Fig. 10.

Necrotic tumor simulation. The top row corresponds to the true fractions. The bottom row corresponds to the estimated fractions using our proposed method. From left to right, each column corresponds to the estimated fractions using our proposed method. From left to right, each column corresponds to tumor, normal tissue, blood and necrotic tissue. The red circle shown in (a) marks the boundary of the tumor ROI used for the single tissue model comparison. The green ellipse in (b) marks the boundary of the normal tissue ROI.

Fig. 11.

Tissue TACs estimated from the metastatic tumor simulation. The error bars show the standard deviation of the estimated TACs. Estimated normal tissue TACs match closely with the true normal tissue TAC for both methods. Note the estimated normal tissue curves overlap with each other.

Fig. 12.

Tissue TACs estimated from the necrotic tumor simulation. The error bars show the standard deviation of the estimated TACs. Estimated normal tissue TACs match closely with the true normal tissue TAC for both methods.

Fig. 13.

(a) CT image of a transverse slice of the patient data. Arrows show selected lesions in the CT image. (b) Estimated tissue TAC using the proposed mixture model. (c) Static PET image reconstructed by summing the time frames from the last 20 minutes of the dynamic PET reconstruction. The blue and green ellipses show the ROIs used for contrast comparison in Table III between tumor and normal tissue region. The area inside the blue ellipse corresponds to the tumor ROI, the green ellipse corresponds to normal tissue ROI. (d) Fusion of the static PET image and the CT image.

Fig. 14. (Left) Estimated tissue fractions from patient data. (Right) Fusion images of the tissue fractions and the CT image. From top to bottom, the fraction components are tumor, normal tissue, blood and zero activity tissue. Arrow shows the descending aorta.

TABLE I

Metastatic phantom tumor volume statistics

Model	Relative bias	Relative std	rmsep
Single tissue	18.9%	2.9%	21.8%
Mixture	-1.0%	2.6%	3.6%

TABLE II

Necrotic phantom phantom tumor volume statistics

Model	Relative bias	Relative std	RMSEP
Single tissue	26.3%	1.2%	27.5%
Mixture	1.2%	2.2%	3.4%

TABLE III

Contrast comparison between tumor and normal tissue ROI means

Model	Tumor	Normal tissue	Contrast
Static PET	23.8	14.6	1.63
Mixture	0.68	0.02	34

TABLE IV

Computational cost of the two stage algorithm

Algorithm component	Time (s)
Estimate B (Eqn. 39)	8.3
Estimate A (Eqn. 47)	43.3
Estimate F (Eqn. 46)	1.4
Estimate n (Eqn. 48)	0.06
First stage total ($N_b = 50$, $Q_{max} = 3$)	10582
Second stage total (20 iterations)	894