Generalized separable parameter space techniques for fitting 1K-5K serial compartment models

Dan J. Kadrmas[a)]
Utah Center for Advanced Imaging Research (UCAIR), Department of Radiology, University of Utah, 729 Arapeen Drive, Salt Lake City, Utah 84108-1218 and MultiFunctional Imaging, 615 Arapeen Drive, Suite 302, Salt Lake City, Utah 84108-1254

M. Bugrahan Oktay
Utah Center for Advanced Imaging Research (UCAIR), Department of Radiology, University of Utah, 729 Arapeen Drive, Salt Lake City, Utah 84108-1218

Purpose: Kinetic modeling is widely used to analyze dynamic imaging data, estimating kinetic parameters that quantify functional or physiologic processes *in vivo*. Typical kinetic models give rise to nonlinear solution equations in multiple dimensions, presenting a complex fitting environment. This work generalizes previously described separable nonlinear least-squares techniques for fitting serial compartment models with up to three tissue compartments and five rate parameters.

Methods: The approach maximally separates the linear and nonlinear aspects of the modeling equations, using a formulation modified from previous basis function methods to avoid a potential mathematical degeneracy. A fast and robust algorithm for solving the linear subproblem with full user-defined constraints is also presented. The generalized separable parameter space technique effectively reduces the dimensionality of the nonlinear fitting problem to one dimension for 2K-3K compartment models, and to two dimensions for 4K-5K models.

Results: Exhaustive search fits, which guarantee identification of the true global minimum fit, required approximately 10 ms for 2K-3K and 1.1 s for 4K-5K models, respectively. The technique is also amenable to fast gradient-descent iterative fitting algorithms, where the reduced dimensionality offers improved convergence properties. The objective function for the separable parameter space nonlinear subproblem was characterized and found to be generally well-behaved with a well-defined global minimum. Separable parameter space fits with the Levenberg-Marquardt algorithm required fewer iterations than comparable fits for conventional model formulations, averaging 1 and 7 ms for 2K-3K and 4K-5K models, respectively. Sensitivity to initial conditions was likewise reduced.

Conclusions: The separable parameter space techniques described herein generalize previously described techniques to encompass 1K-5K compartment models, enable robust solution of the linear subproblem with full user-defined constraints, and are amenable to rapid and robust fitting using iterative gradient-descent type algorithms. © *2013 Author(s). All article content, except where otherwise noted, is licensed under a Creative Commons Attribution 3.0 Unported License.* [http://dx.doi.org/10.1118/1.4810937]

NOMENCLATURE

$b(t)$	Input function
$\breve{b}(t)$	Temporal term (integral of $b(t)$); used in 1K, 3K, and 5K models
$B(t)$	Tracer concentration in whole-blood
f_B	Fractional contribution of $B(t)$ to the imaging measurement
K_i	Net influx macroparameter, $K_i = K_1 k_3 / (k_2 + k_3)$
$K_1, k_2 - k_5$	Rate constants for 1K-5K serial compartment models
$\kappa_B, \kappa_b, \kappa_1, \kappa_2$	Linear parameters of the separable parameter space formulations
$R(t)$	Modeled time-activity curve
$S_1(t; \upsilon_1; b(t))$	Temporal term present in 2K-5K compartment models
$S_2(t; \upsilon_1, \upsilon_2; b(t))$	Temporal term present in 4K-5K compartment models
υ_1, υ_2	Nonlinear parameters of the separable parameter space formulations
V_D	Volume of distribution

I. INTRODUCTION

Dynamic imaging techniques can measure and characterize temporal changes in imaging signals for many modalities, providing more indepth functional information than static images that provide a snapshot or time-average over the
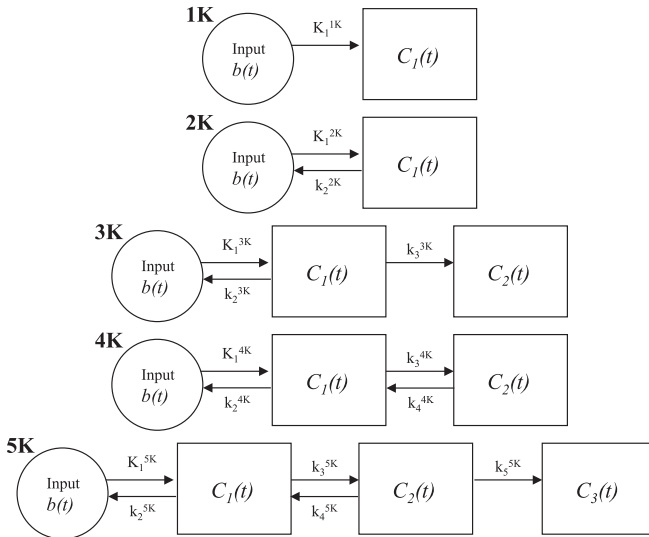
FIG. 1. Generic serial compartment models, each consisting of an input driving 1–3 tissue compartments in series that exchange according to the labeled rate parameters. We use a shorthand nomenclature to quickly reference each generic model as shown. For example, the "3K" model refers to the model with input plus two additional compartments and three rate parameters.

acquisition period. Analysis of dynamic images usually involves applying a kinetic model that describes the temporal behavior of image voxels or regions in terms of a set of unknown scalar parameters. Many types of kinetic models exist, most involving a combination of scalar coefficients that weight temporal terms. These temporal terms are functions of time and may also depend on additional scalar parameters. The archetypal example kinetic model is the compartment model,[1–4] which comprises a series of homogenous compartments driven by an input function, and where temporal exchange between compartments is governed by rate parameters and simple linear differential equations. The solutions to such equations, however, are nonlinear and present a complex fitting environment. Figure 1 presents several generic serial compartment models in order of increasing complexity, along with a shorthand nomenclature that will be used in this paper to quickly reference each generic model. This paper is focused on the mathematical problem of fitting compartment model equations to dynamic imaging data in order to estimate best-fit kinetic rate parameters for a predetermined fitting criterion (e.g., weighted least-squares). Other related issues, such as measurement of the input function or selection of fitting weights, fall outside the scope of this work and are not discussed.

Exchange of tracer between connected compartments is governed by linear differential equations of the form

$$\frac{dC_i}{dt} = \sum k_{in} C_{in} - \left(\sum k_{out} \right) C_i, \tag{1}$$

where the subscripts *in* and *out* refer to rate parameters that describe exchange into (from compartment C_{in}) or out of compartment C_i, respectively. The input function, $b(t)$, drives the system; it is assumed in this work that the input function

is known *a priori* from direct measurement or some other estimation technique. We also assume that all tracer activity concentrations are decay corrected, although the effects of radioactive decay could also easily be included in the modeling equations with a few simple modifications. The imaging measurement $R(t)$ typically cannot measure each compartment individually; rather, the imaging signal comprises the sum over all compartments, $C(t; b(t), \{k_i\})$, often with the addition of a vascular term due to imaging signal from whole-blood, $B(t)$,

$$R(t) = f_B B(t) + (1 - f_B) C(t; b(t), \{k_i\}), \tag{2}$$

where f_B is the fractional contribution of $B(t)$ to the imaging measurement. The rate parameters $\{k_i\}$, along with f_B, comprise the unknown parameters of the model to be estimated. While the differential equations are linear in these unknowns, the solution equations are nonlinear, containing weighted sums of exponentials convolved with the input function. As such, fitting compartment models to measured datasets involves a nonlinear minimization problem.

A large body of work has been performed in addressing the compartment model fitting problem. Perhaps the most robust—but also most computationally demanding—approaches for estimating individual rate parameters for multicompartment models involve nonlinear least-squares estimation (NLLS) in a variety of forms. Numerous nonlinear curve-fitting algorithms have been investigated, including gradient-based algorithms such as Levenberg-Marquardt,[5,6] ridge-regression,[7–11] and simulated annealing methods.[12–15] When using conventional compartment model formulations, the multidimensional fitting space can be large and complex, and it often contains local minima and/or broad shallow valleys that slow or confound iterative minimization routines. The result is sensitive to initial conditions, and one can never be certain that the true global optimum has been reached.

One of the most promising approaches to improving robustness and fitting speed is to reformulate the compartment model solution equations into separable linear and nonlinear components prior to fitting. Separable nonlinear least-square techniques[16–19] can then be used, either directly or indirectly, to simplify and accelerate the fitting problem. This approach has been used for years in fitting 2K compartment models,[9,20–24] and more recently has been extended to 3K and 4K models.[25–28] The nonlinear components of the reformulated equations have largely been interpreted as forming basis functions for the kinetic model, and the fitting algorithm searches among these bases to identify (weighted sums of) curves that best-match the measured data. The best-fit basis functions and linear weighting parameters are then used to compute corresponding kinetic rate parameters (and/or macroparameters) of the compartment model. We collectively refer to these approaches as basis function methods throughout this paper.

Basis function methods bring the advantage of simplifying the nonlinear compartment model fitting problem by effectively separating the nonlinear aspects of the fit (i.e., identification of the best-fit basis functions) from the linear

aspects (i.e., finding best-fit scalar weights for a given set of basis functions). However, current basis function methods have a few limitations. The basis function formulations for 4K models exhibit a potential degeneracy in the nonlinear parameters of the fit (see Sec. II.A below), which can lead to indeterminate results. Similarly, full implementation of boundary conditions and user-definable constraints on kinetic rate parameter estimates has not yet been described. Moreover, the algorithms employed for identifying the best-fit basis functions have largely been based on exhaustive search or simplex strategies, and fast iterative gradient-descent algorithms have not been broadly applied in this context.

In this work, we describe an alternate mathematical formulation of basis function methods that generalizes separable parameter space techniques for serial compartment models with up to three tissue compartments and five rate parameters. The alternate formulation avoids a potential degeneracy that may arise for previously described basis function formulations. A theoretical perspective of interpreting the resulting equations as effectively reducing the dimensionality of the solution space of the fitting problem is then discussed, followed by a description of a rigorous method for solving the linear subproblem with full user-defined constraints. The reformulated modeling equations are validated against conventional model formulations, and we demonstrate that both formulations provide identical time-activity curves and have the same least-squares global minimum. The application of fast iterative gradient descent algorithms for solving the nonlinear subproblem is then explored by characterizing separable parameter space objective functions. Finally, the performance of both exhaustive search and the iterative Levenberg-Marquardt algorithms are studied for separable parameter space fits and compared to fits using the conventional model formulations.

II. THEORY

Without loss of generality, the theory for the generalized separable parameter space modeling formulation is presented in the context of kinetic modeling in dynamic nuclear medicine imaging, e.g., positron emission tomography (PET) and single-photon emission computed tomography (SPECT). As such, quantities are described as tracer radioactivity concentrations. The concepts and mathematics, however, are generalizable to other dynamic imaging modalities and nonimaging applications of kinetic modeling.

II.A. Generalized kinetic modeling formulation

Let $\kappa_B, \kappa_b, \kappa_1, \kappa_2, \upsilon_1$, and υ_2 be scalar parameters that will be further described below, and let $B(t)$ be the whole-blood activity concentration. Further, define the following temporal terms that depend on the input function, $b(t)$, as well as scalar parameters υ_1 and υ_2,

$$\breve{b}(t) \equiv \int_0^t b(\tau) d\tau, \tag{3}$$

$$S_1(t; \upsilon_1; b(t)) \equiv \int_0^t e^{-\upsilon_1(t-\tau)} b(\tau) d\tau, \tag{4}$$

$$S_2(t; \upsilon_1, \upsilon_2; b(t)) \equiv \int_0^t (e^{-\upsilon_2(t-\tau)} - e^{-\upsilon_1(t-\tau)}) b(\tau) d\tau. \tag{5}$$

Note that the temporal terms $S_1(t; \upsilon_1; b(t))$ and $S_2(t; \upsilon_1, \upsilon_2; b(t))$ are convolutions of exponentials with the input function. We can now write a generalized modeling equation

$$R(t) = \kappa_B B(t) + \kappa_b \breve{b}(t) + \kappa_1 S_1(t; \upsilon_1; b(t))$$
$$+ \kappa_2 S_2(t; \upsilon_1, \upsilon_2; b(t)), \tag{6}$$

which is linear in $\{\kappa_B, \kappa_b, \kappa_1, \kappa_2\}$ and nonlinear in $\{\upsilon_1, \upsilon_2\}$. Defining $\kappa_B, \kappa_b, \kappa_1, \kappa_2, \upsilon_1$, and υ_2 as listed in Tables I and II, it can easily be shown that this general formulation encompasses the solution equations for well known 1K-4K serial compartment models,[1,2,29,30] as well as the lesser-used 5K serial compartment model.[31]

We refer to Eq. (6) as the "generalized separable parameter space formulation." For 2K models, the formulation presented above is equivalent to that previously used with the application of separable nonlinear least-squares for these models.[20] Similarly, for 3K models it is identical to the basis function for plasma input compartment (BAFPIC) method of Hong et al.[28] The proposed generalized formulation extends the approach to include the 5K compartment model, and for $\geq$4K models, there is a subtle difference as compared to previously described basis function methods.[27] The difference lies in our definition of the S_2 term, which in previous approaches was defined as $S_2(t; \upsilon_2; b(t)) \equiv \int_0^t e^{-\upsilon_2(t-\tau)} b(\tau) d\tau$. Subtracting $e^{-\upsilon_1 t}$ in the S_2 term as in Eq. (5) ensures that S_1 and S_2 are determinant and distinguishable (i.e., $S_1 \neq S_2$ for all nonzero values of $\upsilon_{1,2}$), which avoids the potential degeneracy in the previous approach that could arise for $\upsilon_1 \approx \upsilon_2$. Further, our formulation of S_2 immediately leads to the constraint $0 < \upsilon_2 < \upsilon_1$ in order to ensure positivity for this temporal term, which is required in order to limit Eq. (6) to represent valid compartment model solutions. Similarly, the special case of $\upsilon_1 = \upsilon_2$ in our formulation implies that both k_4 and k_5 must be zero in order to represent a valid compartment model, which corresponds to simplification to a 3K model (see below). Analogous constraints and boundary conditions could be applied when using previously described basis function formulations,[27] for example, by implementing a constraint to ensure that υ_1 and υ_2 differ by some minimum value. However, we find that the description and interpretation of these cases may be more straightforward when using the formulation described in Eq. (5).

The constraints listed in Tables I and II limit the generalized formulation to valid compartment model solutions with positivity constraints on the conventional kinetic rate parameters $K_1, k_2 - k_5$. For 4K and 5K models, the constraint $0 < \upsilon_2 < \upsilon_1$ is equivalent to requiring k_2 to be positive and k_3 and k_4 to be non-negative, and $0 < \kappa_2 < \kappa_1$ ensures positivity for $k_2 - k_4$. Of note, we have used positivity constraints here (except for the vascular term κ_B) rather than non-negativity constraints, thus the constraints preclude solutions with rate

TABLE I. Nonlinear parameter definitions for 1K-5K compartment models.

Model	v_1	v_2	Constraint
1K	0	0	n/a
2K	k_2	0	$0 < v_1$
3K	$k_2 + k_3$	0	$0 < v_1$
4K	$\frac{1}{2}\left\{(k_2+k_3+k_4)+\sqrt{(k_2+k_3+k_4)^2-4k_2k_4}\right\}$	$\frac{1}{2}\left\{(k_2+k_3+k_4)-\sqrt{(k_2+k_3+k_4)^2-4k_2k_4}\right\}$	$0 < v_2 < v_1$
5K	$\frac{1}{2}\left\{\begin{array}{l}(k_2+k_3+k_4+k_5)+\\ \sqrt{(k_2+k_3+k_4+k_5)^2-4(k_2k_4+k_2k_5+k_3k_5)}\end{array}\right\}$	$\frac{1}{2}\left\{\begin{array}{l}(k_2+k_3+k_4+k_5)-\\ \sqrt{(k_2+k_3+k_4+k_5)^2-4(k_2k_4+k_2k_5+k_3k_5)}\end{array}\right\}$	$0 < v_2 < v_1$

parameters equal to zero. Solutions where a rate parameter is zero are encompassed via model simplification, e.g., a 4K model with $k_4 = 0$ simplifies to a 3K model. This approach also ensures that the modeling formulation does not attempt to compute inestimable rate parameters. For example, if $k_3 = 0$, then no information is available regarding k_4 and k_5 (as there is no tracer exchange with the unobservable compartment), and the estimable model reduces to a 2K model.

II.B. Construction of the WSSE objective function

For dynamic imaging, the measured data are generally discrete samples in time. Depending on the modality and other factors, two discretization schemes are commonly encountered—instantaneous samples, and samples averaged over time-frame durations. For each case we define the following discretizations, where $j = 1 \ldots M$ represents the time (time-frame) index:

Instantaneous samples:

$$R_j \equiv R(t_j) \quad B_j \equiv B(t_j) \quad \breve{b}_j \equiv \breve{b}(t_j)$$

$$S_{1,j}(v_1;b) \equiv S_1(t_j;v_1;b) \quad S_{2,j}(v_1,v_2;b) \equiv S_2(t_j;v_1,v_2;b).$$
$$\tag{7}$$

Averaged over time-frame durations:

$$R_j \equiv \frac{1}{(t_{j2}-t_{j1})}\int_{t_{j1}}^{t_{j2}} R(\tau)d\tau$$

$$B_j \equiv \frac{1}{(t_{j2}-t_{j1})}\int_{t_{j1}}^{t_{j2}} B(\tau)d\tau$$

$$\breve{b}_j \equiv \frac{1}{(t_{j2}-t_{j1})}\int_{t_{j1}}^{t_{j2}} \breve{b}(\tau)d\tau$$

$$S_{1,j}(v_1;b) \equiv \frac{1}{(t_{j2}-t_{j1})}\int_{t_{j1}}^{t_{j2}} S_1(\tau;v_1;b)d\tau$$

$$S_{2,j}(v_1,v_2;b) \equiv \frac{1}{(t_{j2}-t_{j1})}\int_{t_{j1}}^{t_{j2}} S_2(\tau;v_1,v_2;b)d\tau. \tag{8}$$

In addition, we distinguish between modeled values $\hat{R}_j$ and noisy measurements $\tilde{R}_j$ using a carat and tilde, respectively.

Using either discretization, the generalized kinetic model at each time sample j is written

$$\hat{R}_j = \kappa_B B_j + \kappa_b \breve{b}_j + \kappa_1 S_{1,j}(v_1;b) + \kappa_2 S_{2,j}(v_1,v_2;b). \tag{9}$$

The model fitting problem amounts to finding the values of the parameters $\kappa_B, \kappa_b, \kappa_1, \kappa_2, v_1,$ and v_2 that minimize some objective function. We consider the weighted least-squares (WLS) criterion, and write the weighted sum-squared error objective function as

$$WSSE = \sum_{j=1}^{T} w_j(\hat{R}_j - \tilde{R}_j)^2$$

$$= \sum_{j=1}^{T} w_j[\kappa_B B_j + \kappa_b \breve{b}_j + \kappa_1 S_{1,j}(v_1;b)$$

$$+ \kappa_2 S_{2,j}(v_1,v_2;b) - \tilde{R}_j]^2, \tag{10}$$

where w_j are the weights for each time sample j.

TABLE II. Linear parameter definitions for 1K-5K compartment models.

Model	κ_B	κ_b	κ_1	κ_2	Constraints
1K	f_B	$(1-f_B)K_1$	0	0	$0 \le \kappa_B \le 1; 0 < \kappa_b$
2K	f_B	0	$(1-f_B)K_1$	0	$0 \le \kappa_B \le 1; 0 < \kappa_1$
3K	f_B	$\frac{(1-f_B)K_1k_3}{k_2+k_3}$	$\frac{(1-f_B)K_1k_2}{k_2+k_3}$	0	$0 \le \kappa_B \le 1; 0 < \kappa_b; 0 < \kappa_1$
4K	f_B	0	$(1-f_B)K_1$	$\frac{(1-f_B)K_1}{(v_2-v_1)}(v_2-k_3-k_4)$	$0 \le \kappa_B \le 1; 0 < \kappa_2 < \kappa_1$
5K	f_B	$\frac{(1-f_B)K_1k_3k_5}{k_2k_4+k_2k_5+k_3k_5}$	$\frac{(1-f_B)K_1k_2(k_4+k_5)}{k_2k_4+k_2k_5+k_3k_5}$	$\frac{(1-f_B)K_1}{(v_2-v_2)}\left[(k_3+k_4+k_5-v_2)-\frac{k_3k_5}{v_2}\right]$	$0 \le \kappa_B \le 1; 0 < \kappa_b; 0 < \kappa_2 < \kappa_1$

II.C. Application of separable nonlinear least-squares

The generalized formulation of Eq. (6) has 6 degrees-of-freedom (κ_B, κ_b, κ_1, κ_2, υ_1, υ_2), as does the conventional model formulation for the 5K model (f_B, K_1, $k_2 - k_5$); however, the generalized formulation is written to be explicitly linear in κ_B, κ_b, κ_1, and κ_2, and nonlinear in υ_1 and υ_2 (which appear as exponents in the convolutions of temporal terms S_1 and S_2). Inspection of the 2K and 3K models (see, e.g., Tables I and II) reveals that there is inherently one convolution integral containing a single free parameter in the exponent (υ_1), resulting in one nonlinear degree-of-freedom; the remaining degrees-of-freedom are linear. Likewise, inspection of the 4K and 5K models reveals two inherent convolution integrals with differing exponents. The free parameters in the exponents, υ_1 and υ_2, are independent parameters: υ_1 cannot be written in terms of (κ_B, κ_b, κ_1, κ_2, υ_2), and likewise υ_2 cannot be written in terms of (κ_B, κ_b, κ_1, κ_2, υ_1). As such, two separate free parameters are required to represent these convolution integrals, and these models inherently have two nonlinear degrees-of-freedom. Since the generalized formulation of Eq. (6) is written explicitly to have the minimum number of nonlinear free parameters (one for 2K-3K models, two for 4K-5K models), and likewise the maximum number of linear free parameters (the κs), this formulation can be considered to maximally separate the linear and nonlinear parameters.

Separating the linear and nonlinear parameters in this manner provides for the full benefit of separable nonlinear least-squares techniques.[16–19] The fitting problem is essentially separated into two problems: a nonlinear fit in (υ_1, υ_2), coupled simultaneously with a linear fit in (κ_B, κ_b, κ_1, κ_2). Previous basis function methods have largely used exhaustive search or downhill simplex algorithms to search among the (υ_1, υ_2) "basis functions," analytically solving the linear subproblem to find the best-fit (κ_B, κ_b, κ_1, κ_2) for each point (υ_1, υ_2) encountered along the way. In this section, we show that the linear subproblem can also be directly incorporated into the nonlinear fit, providing a "Reduced Parameter Space Reformulation" with interesting theoretical perspectives that illustrate the value of separable parameter space fitting techniques.

Consider the linear subproblem that arises at each iteration/step of the nonlinear fit: finding (κ_B, κ_b, κ_1, κ_2) that minimizes *WSSE* given the current value of (υ_1, υ_2). For the moment, we ignore constraints on the κs and consider the unconstrained linear subproblem. Taking the derivative of *WSSE* [Eq. (10)] with respect to each linear parameter and setting the result to zero, we obtain the linear subproblem

$$\mathbf{A}\mathbf{x} = \mathbf{b},$$

where

$$\mathbf{A} = \begin{pmatrix} \sum_{j=1}^{T} w_j B_j^2 & \sum_{j=1}^{T} w_j B_j \breve{b}_j & \sum_{j=1}^{T} w_j S_{1,j}(\upsilon_1; b) B_j & \sum_{j=1}^{T} w_j S_{2,j}(\upsilon_1, \upsilon_2; b) B_j \\ \cdot & \sum_{j=1}^{T} w_j \breve{b}_j^2 & \sum_{j=1}^{T} w_j S_{1,j}(\upsilon_1; b) \breve{b}_j & \sum_{j=1}^{T} w_j S_{2,j}(\upsilon_1, \upsilon_2; b) \breve{b}_j \\ \cdot & \cdot & \sum_{j=1}^{T} w_j S_{1,j}^2(\upsilon_1; b) & \sum_{j=1}^{T} w_j S_{1,j}(\upsilon_1; b) S_{2,j}(\upsilon_1, \upsilon_2; b) \\ \cdot & \cdot & \cdot & \sum_{j=1}^{T} w_j S_{2,j}^2(\upsilon_1, \upsilon_2; b) \end{pmatrix}$$

$$\mathbf{x} = \begin{pmatrix} \kappa_B \\ \kappa_b \\ \kappa_1 \\ \kappa_2 \end{pmatrix}, \qquad \mathbf{b} = \begin{pmatrix} \sum_{j=1}^{T} w_j \tilde{R}_j B_j \\ \sum_{j=1}^{T} w_j \tilde{R}_j \breve{b}_j \\ \sum_{j=1}^{T} w_j \tilde{R}_j S_{1,j}(\upsilon_1; b) \\ \sum_{j=1}^{T} w_j \tilde{R}_j S_{2,j}(\upsilon_1, \upsilon_2; b) \end{pmatrix}, \tag{11}$$

and "." denotes symmetric matrix elements.

The conventional approach, as used with previously described basis function methods, is to analytically invert the linear subproblem ($\hat{\mathbf{x}} = \mathbf{A}^{-1}\mathbf{b}$) to directly compute the best-fit κs that minimize $WSSE$ given the current values of υ_1 and υ_2. The estimated κs are then fed back to the nonlinear fitting algorithm, and the fit in (υ_1, υ_2) progresses accordingly.

Consider an alternative approach, however, where the inverted linear subproblem (which computes best-fit κs as functions of υ_1, υ_2, and measured/known quantities) is substituted back into the original modeling Eq. (6). The result is a reformulated modeling equation, which we call the Unconstrained Generalized Reduced Parameter Space Reformulation, which effectively removes the linear parameters from the model:

Unconstrained generalized reduced parameter space reformulation:

$$
\hat{R}_j = \left(\left(\begin{pmatrix} \sum_{j=1}^{T} w_j B_j^2 & \sum_{j=1}^{T} w_j B_j \breve{b}_j & \sum_{j=1}^{T} w_j S_{1,j}(\upsilon_1;b) B_j & \sum_{j=1}^{T} w_j S_{2,j}(\upsilon_1, \upsilon_2;b) B_j \\ . & \sum_{j=1}^{T} w_j \breve{b}_j^2 & \sum_{j=1}^{T} w_j S_{1,j}(\upsilon_1;b) \breve{b}_j & \sum_{j=1}^{T} w_j S_{2,j}(\upsilon_1, \upsilon_2;b) \breve{b}_j \\ . & . & \sum_{j=1}^{T} w_j S_{1,j}^2(\upsilon_1;b) & \sum_{j=1}^{T} w_j S_{1,j}(\upsilon_1;b) S_{2,j}(\upsilon_1, \upsilon_2;b) \\ . & . & . & \sum_{j=1}^{T} w_j S_{2,j}^2(\upsilon_1, \upsilon_2;b) \end{pmatrix}^{-1} \begin{pmatrix} \sum_{j=1}^{T} w_j \tilde{R}_j B_j \\ \sum_{j=1}^{T} w_j \tilde{R}_j \breve{b}_j \\ \sum_{j=1}^{T} w_j \tilde{R}_j S_{1,j}(\upsilon_1;b) \\ \sum_{j=1}^{T} w_j \tilde{R}_j S_{2,j}(\upsilon_1, \upsilon_2;b) \end{pmatrix} \right)^T \right.
$$
$$
\left. \times \begin{pmatrix} B_j \\ \breve{b}_j \\ S_{1,j}(\upsilon_1;b) \\ S_{2,j}(\upsilon_1, \upsilon_2;b) \end{pmatrix} \right).
\tag{12}
$$

In essence, the reformulated model equation has been constrained to only permit solutions that minimize the $WSSE$ objective function *in the linear sense*. Inspection of each term in Eq. (12) reveals that the reformulated model depends on υ_1, υ_2, and the measurements; however, no linear free parameters (κs) remain. One interpretation of this is that the application of separable nonlinear least-squares can be considered to constrain the solution space to only include solutions that minimize the objective function in the linear sense, effectively reducing the dimensionality of the fit to only the nonlinear free parameters. More specifically, the dimensionality of the "reduced" fitting space for 1K-5K serial compartment models becomes:

Model	Conventional formulation	Reduced parameter space reformulation
1K	2D (f_B, K_1)	0D (no free parameters)
2K	3D (f_B, K_1, k_2)	1D (υ_1)
3K	4D ($f_B, K_1, k_2 - k_3$)	1D (υ_1)
4K	5D ($f_B, K_1, k_2 - k_4$)	2D (υ_1, υ_2)
5K	6D ($f_B, K_1, k_2 - k_5$)	2D (υ_1, υ_2)

While of limited practical interest (due to difficulty in constraining the κs with this approach), this theoretical perspective illustrates the value of separable parameter space fitting techniques (including basis function methods). The effective dimensionality of the nonlinear fitting problem is greatly reduced, simplifying and speeding the fit. Of note, any iterative minimization procedure employed to solve the nonlinear subproblem would act effectively the same regardless of whether it uses Eq. (11) or (12), except for approaches which implement constraints on the κs.

II.D. Constraining fitted parameters in the linear subproblem

The constraints listed in Tables I and II ensure that the generalized separable parameter space formulation [Eq. (6)] represents valid compartment models with positivity constraints on all kinetic parameters (plus $0 \leq f_B \leq 1$). In practice, one may well wish to impose additional constraints on parameter estimates for the particular application at hand. In either case, constraints on the nonlinear parameters (υ_1, υ_2) can be imposed in the standard fashion for whichever nonlinear fitting algorithm is employed. Implementing constraints on the

linear subproblem, however, is less straightforward. Direct inversion of the linear subproblem of Eq. (11) does not account for any constraints on the κs, and the direct analytical solution to the linear subproblem is unconstrained. *Ad hoc* approaches to implementing a certain level of constraint for the linear subproblem have been proposed;[32] however, these approaches do not guarantee that the resulting fitted parameters minimize the objective function within the parameter ranges allowed.

The Appendix describes a pseudo-code algorithm for solving the linear subproblem with full user-defined constraints on the estimated parameter values (κs). The algorithm first computes the direct inversion solution. If any parameter is out of range, then an exhaustive search over that parameter is performed, analytically computing the direct inverse for the remaining parameters at each step. When more than one parameter goes out of range, nested loops are used to exhaustively search among those parameters whose analytical solution is out of range while computing analytical solutions for those parameters remaining in range. While this algorithm can require a large number of steps (i.e., when exhaustive searches are required for more than one κ), each step is extremely fast as no new convolution integrals need be performed and many of the sums can be reused for each step. The resulting computational cost per step is orders-of-magnitude lower than the cost per step (or iteration) of the nonlinear subproblem [where changing υ_1 and/or υ_2 requires recalculation of convolution integral(s)].

One may also desire to apply physiologic equality constraints to certain parameters, e.g., setting K_1/k_2 to some fixed value known *a priori*. A number of different such scenarios could arise for different tracers and modeling applications, and discussion of each falls outside the scope of this work. However, separable parameter space fitting approaches could still be used by applying the equality constraint to the original modeling equations and rederiving the separable parameter space formulation. Depending on the equality constraint(s) applied, the resultant formulations may have fewer linear and/or nonlinear free parameters than shown in this work for the generalized formulations.

II.E. Fitting algorithms for the nonlinear subproblem

Two fitting algorithms for solving the nonlinear subproblem were used in this work. The first algorithm, exhaustive search, is equivalent to the basis function search algorithms used with much of the prior work on basis function modeling methods (with some slight differences in the details of implementation). In some sense, this algorithm can be considered as searching among the temporal terms ("basis functions") that correspond to each nonlinear model parameter (υ_1, υ_2). Taking a somewhat different perspective, one can also treat the nonlinear parameters themselves as parameters of the fit and apply iterative gradient-descent fitting approaches. In this work, we study the Levenberg-Marquardt algorithm for this purpose.

II.E.1. Exhaustive search algorithm

The dimensionality of the separable parameter space nonlinear subproblem—only one or two free parameters for 2K-3K and 4K-5K compartment models, respectively—is small enough that the entire solution space (within appropriate parameter ranges) can be exhaustively searched to an arbitrary precision with a reasonable amount of computational effort. Such brute-force exhaustive search *guarantees* identification of the global minimum to within the selected search precision and parameter ranges. As such, exhaustive search provides the most robust fit for the given data, weights, and input functions. In this work, we implemented exhaustive search routines in C for solving the nonlinear subproblem (finding best-fit υ_1 and υ_2). The routine loops over υ_1 from 0.0 to 2.0 min^{-1} in 1000 steps, sampling υ_1 at a precision of 0.002 min^{-1}. Here, we chose uniformly spaced sampling in order to simplify discussion of the precision of the algorithm; for practical application, logarithmic spacing would likely be more appropriate as used in most previous implementations of these algorithms. For the 4K and 5K models, a nested loop over υ_2 was likewise included for each value of υ_1. Recalling that υ_2 should be constrained less than υ_1 in order to represent a valid compartment model, this inner loop ranged from 0.0 up to the current value of υ_1 in steps of 0.002 min^{-1}. This resulted in 1000 total steps for the 2K-3K models, and $\sim$500 000 steps for the 4K-5K models. For comparison, consider that exhaustive search with conventional compartment model formulations would require 10^9, 10^{12}, and 10^{15} steps for 2K, 3K, and 4K models, respectively—becoming computationally intractable.

II.E.2. Levenberg-Marquardt algorithm

Nonlinear minimization with the Levenberg-Marquardt algorithm was also implemented in C for both the generalized separable parameter space formulation and for conventional 1K-5K compartment model formulations. Single-threaded coding was used with no parallelization. Close attention was paid to code optimization for both model formulations, and all concurrent routines applicable to both algorithms were shared in order to match the implementations as closely as possible and to ensure valid comparisons between the two implementations could be performed. For the conventional implementation, the vascular fraction (f_B) and all applicable rate parameters ($K_1, k_2 - k_5$) were estimated by the fitting algorithm, whereas only the nonlinear parameters υ_1, υ_2 were fit for the separable parameter space implementation. The algorithm was implemented as described in Press *et al.*,[33] where the second derivative term of the Hessian matrix was ignored. Iterations were stopped after a successful update when the fractional change in *WSSE* reached 0.0001 or less, indicating convergence to a (possibly local) minimum. The magnitude of the update vector was bound to fall within the range [0.01,1.00] in order to avoid wild steps as well as to reduce the number of steps required to traverse a shallow valley of complicated topology. Various initial conditions were used for the evaluation studies as described in Sec. IV.

TABLE III. Recovery of kinetic rate parameters.

Model	f_B	K_1	k_2	k_3	k_4	k_5
1K	κ_B	$\frac{\kappa_1}{(1-f_B)}$	0	0	0	0
2K	κ_B	$\frac{\kappa_1}{(1-f_B)}$	υ_1	0	0	0
3K	κ_B	$\frac{\kappa_b+\kappa_1}{(1-f_B)}$	$\frac{\kappa_1\upsilon_1}{\kappa_b+\kappa_1}$	$\frac{\kappa_b\upsilon_1}{\kappa_b+\kappa_1}$	0	0
4K	κ_B	$\frac{\kappa_1}{(1-f_B)}$	$\frac{(\kappa_1-\kappa_2)\upsilon_1+\kappa_2\upsilon_2}{\kappa_1}$	$\frac{\kappa_2(\kappa_1-\kappa_2)(\upsilon_2-\upsilon_1)^2}{\kappa_1[(\kappa_1-\kappa_2)\upsilon_1+\kappa_2\upsilon_2]}$	$\frac{\kappa_1\upsilon_1\upsilon_2}{(\kappa_1-\kappa_2)\upsilon_1+\kappa_2\upsilon_2}$	0
5K	κ_B	$\frac{\kappa_b+\kappa_1}{(1-f_B)}$	$\frac{(\kappa_1-\kappa_2)\upsilon_1+\kappa_2\upsilon_2}{\kappa_b+\kappa_1}$	$\left\{\begin{array}{c}\frac{\kappa_2(\upsilon_1-\upsilon_2)+\kappa_b\upsilon_1}{(\kappa_b+\kappa_1)}+\\[4pt]\frac{\kappa_2\upsilon_2(\upsilon_2-\upsilon_1)}{(\kappa_1-\kappa_2)\upsilon_1+\kappa_2\upsilon_2}\end{array}\right\}$	$\left\{\begin{array}{c}\frac{\kappa_1\upsilon_1\upsilon_2}{(\kappa_1-\kappa_2)\upsilon_1+\kappa_2\upsilon_2}-\\[4pt]\frac{\kappa_b\upsilon_1\upsilon_2}{k_3(\kappa_b+\kappa_1)}\end{array}\right\}$	$\frac{\kappa_b\upsilon_1\upsilon_2}{k_3(\kappa_b+\kappa_1)}$

II.F. Recovery of individual rate parameters K_1-k_5

After completing the fit, the kinetic rate parameters K_1, $k_2 - k_5$, as well as f_B, are easily calculated from the best-fit parameters $\kappa_B, \kappa_b, \kappa_1, \kappa_2, \upsilon_1$, and υ_2. The calculation does merit some discussion of parameter estimability, however. Table III provides conversions for the 1K-5K serial compartment models. In all cases $f_B = \kappa_B$. When macro parameters such as the volume of distribution or net influx are desired, they can either be computed using the individual rate parameters, or more directly in some cases (e.g., the net influx parameter for the 3K model is more directly calculated as $K_i = \kappa_b/(1 - f_B)$).

The above equations provide one-to-one correspondence between any set of separable parameter space variables and a set of conventional rate parameters, provided that divide-by-zero singularities are not encountered. The majority of such potential singularities are avoided through application of the constraints listed in Tables I and II, which concurrently ensure that the generalized modeling equations represent valid 1K-5K compartment models. The remaining potential singularities are avoided through model simplification (as described in Sec. II.A). For example, consider the potential singularity of $f_B = 1.0$, affecting all K_1 calculations. If the best-fit value of f_B is 1.0, this means the best-fit time-activity curve is pure whole-blood and there is no extravascular tissue uptake. In that case, K_1 is zero and all higher order rate parameters ($k_2 - k_5$) are undefined and not meaningful (one could consider this a "0K" model).

III. EVALUATION METHODS

Separable parameter space modeling techniques, which are generalized in this work, have already been thoroughly evaluated in the research literature for both 2K models[9,20–24] and more complex models with a basis function methods.[25–28] As such, the evaluation provided in this work is designed to first broadly characterize separable parameter space model formulations vs conventional compartment model formulations, and then to evaluate the applicability and receptiveness of the generalized separable parameter space formulations to fast gradient-descent fitting algorithms. Sections III.A–III.D describe the evaluations performed in this work. The emphasis of this evaluation is focused on 2K, 3K, and 4K serial compartment models, as these represent the most commonly used compartment models in dynamic PET imaging and cover both 1D (3K) and 2D (4K) separable parameter space solution equations.

III.A. Simulation of time-activity curves

A set of 24 typical input functions and whole-blood curves were selected from dynamic PET exams using ^{15}O-water (2K model), ^{18}F-fluorodeoxyglucose (FDG; 3K model), and ^{11}C-acetate (4K model) in patients with malignant solid tumors at our institution. Populations of kinetic parameters were then generated with a pseudorandom number generator to be randomly distributed within the following ranges: $0.001 \le f_B \le 0.2$, $0.02 \le K_1 \le 0.5$, $0.02 \le k_2 \le 0.7$, $0.002 \le k_3 \le 0.4$, and $0.002 \le k_4 \le 0.4$. Here, each parameter was independently and uniformly distributed over the given range except for k_2, which was set according to a uniform distribution of $0.7 \le K_1/k_2 \le 1.1$. These distributions were not designed to represent any particular tracer or target organ, but rather were set to cover a broad range of values that might be encountered in a variety of applications. The input functions and rate parameter ranges were used to simulate populations of 200–1000 time-activity curves each for the 2K, 3K, and 4K models using the conventional model formulations. The number of curves in each population was heuristically selected to include enough datapoints such that "outlier" points (representing best-fit differences between the separable parameter space and conventional model fits—see, e.g., results Figs. 6 and 7) presented themselves for study. Progressive temporal sampling schemes were used, for example, 18×10, 6×20, 4×30, 5×60, 5×120, 4×300, 3×600 s for a total duration of 72 m for the ^{18}F 3K model simulations.

Noise-free curves were first simulated for each case. Statistical noise was then added using a noise model that approximates the noise characteristics of iteratively reconstructed PET images. Here, Gaussian deviates were simulated for each datapoint with variance inversely proportional to the time-frame durations, and scaled to five different noise levels. When a noisy sample went negative, however, the absolute value was used—providing an asymmetric non-negative distribution roughly approximating the statistical distribution of iteratively reconstructed PET values. The resulting noisy curves had Gaussian-like non-negative statistics that

varied with time-frame duration, providing more realistic test cases than simulating curves that were consistent with either Gaussian or Poisson noise. The noise model did not account for other factors, such as differences due to radioactive decay, deadtime, randoms rates, or reconstruction algorithm, that would be encountered in practice. Note that this noise model is inconsistent with the L2-norm criterion typically used in model fitting (which essentially assumes a Gaussian noise model); however, such inconsistency is a real and present challenge for actual dynamic PET data. Five different noise levels were used, covering the approximate range encountered in the dynamic PET datasets—ranging from high noise in individual voxels to lower noise curves representative of larger tumor ROIs.

III.B. Characterization of separable parameter space WSSE objective functions

Given the reduction in the number of free parameters offered by the separable parameter space formulations, the objective functions reside within 1D (2K-3K models) or 2D (4K-5K models) solution spaces. Thus, they can be plotted and analyzed much more easily than conventional objective functions in three to six dimensions. Using the exhaustive search routines just described, the *WSSE* objective functions were characterized for a number of noisy 3K-5K time-activity curves by storing and plotting the *WSSE* at each value of v_1 (3K) or v_1, v_2 (4K-5K). The resulting objective functions were analyzed for the presence/absence of local minima or other confounding structures that could affect the performance of curve fits using the separable parameter space formulations with various minimization algorithms.

III.C. Characterization of global minimums for both conventional and separable parameter space formulations

In order for the separable parameter space formulations to be a viable alternative for kinetic modeling, the reformulated modeling equations must provide the same global minimum and best-fit rate parameters as the conventional model equations. In order to perform this study, fits to both model formulations need to correctly reach the true global minimums. Identification of the true global minimum is not straightforward, however, as commonly used fitting algorithms may be sensitive to initial conditions and converge to local minima (see Sec. IV.D, for example, data regarding convergence for the Levenberg-Marquardt algorithm for conventional model fits). To overcome these shortcomings, the simulated annealing algorithm[13–15,33] was selected for the conventional model fits. As implemented, the simulated annealing algorithm was independent of starting conditions (aside from the set parameter ranges used for all fits in this paper) and included means for escaping local minima and progressing toward the true global minimum provided that enough iterations and a slow enough relaxation schedule were used. We progressively applied simulated annealing with 10^4–10^8 iterations, increasing

the number of iterations by $10\times$ each step and observing the number of cases in each population where the best-fit solution did not change versus those that continued moving toward a better fit. While this approach does not guarantee that the true global minimum was reached for *all* fits (only exhaustive search could so guarantee), it did allow visualization of the progression of the population of fits toward the true global minima as will be shown in Figs. 6 and 7 below. As such, the simulated annealing algorithm was used with the conventional model formulations in order to identify the true global minimum fits for the simulated population studies. The exhaustive search algorithm was used to identify the global minimum fits for the separable parameter space formulations, where the reduced dimensionality of the separable parameter space nonlinear subproblem makes such exhaustive search computationally feasible (guaranteeing identification of the true global minimum to within the search range and precision for all cases).

III.D. Performance of Levenberg-Marquardt algorithm

The performance of the Levenberg-Marquardt algorithm for solving the nonlinear subproblem of the separable parameter space formulations was studied and compared to fitting performance of the same algorithm for conventional compartment model formulations. Iterative fitting performance was evaluated in terms of convergence to the true global minimum (as opposed to local minima), sensitivity to initial values, and number of iterations and fitting time required to reach convergence. Here the populations of noisy time-activity curves for 2K-4K models described above were fit again using the Levenberg-Marquardt algorithm with both model formulations. Three sets of fits were performed with differing initial conditions. First, each curve was fit repeatedly using 100 randomly generated sets of rate parameters as initial estimates. These fits provide a measure of sensitivity of the iterative fits to the initial parameters. In practice, the initial parameter estimates for conventional model Levenberg-Marquardt fits would be selected manually by the user to be a reasonably close match to the measured curve. In addition to introducing a level of subjectivity into the final result, the manual intervention adds significantly to the real-world fitting time. Rather than replicating this subjective process, randomly selected initial parameters were studied in order to characterize the sensitivity of the fits to these initial conditions. In the second set of fits, the initial parameter values were each set to the minimum value (e.g., 0.001), which results in the Levenberg-Marquardt algorithm finding the (possibly local) minimum closest to the minimum parameter values. Finally, in the third set of fits, the initial parameter values were set to midrange values, defined as being 1/4th the way between the minimum and maximum value for each parameter (e.g., for a parameter ranging from 0.0 to 2.0, the initial value was set at 0.5). The number of iterations and CPU times were recorded for each fit, and the best-fit results were analyzed relative to the true global minimum fits for each case (determined as described in Sec. III.C above).
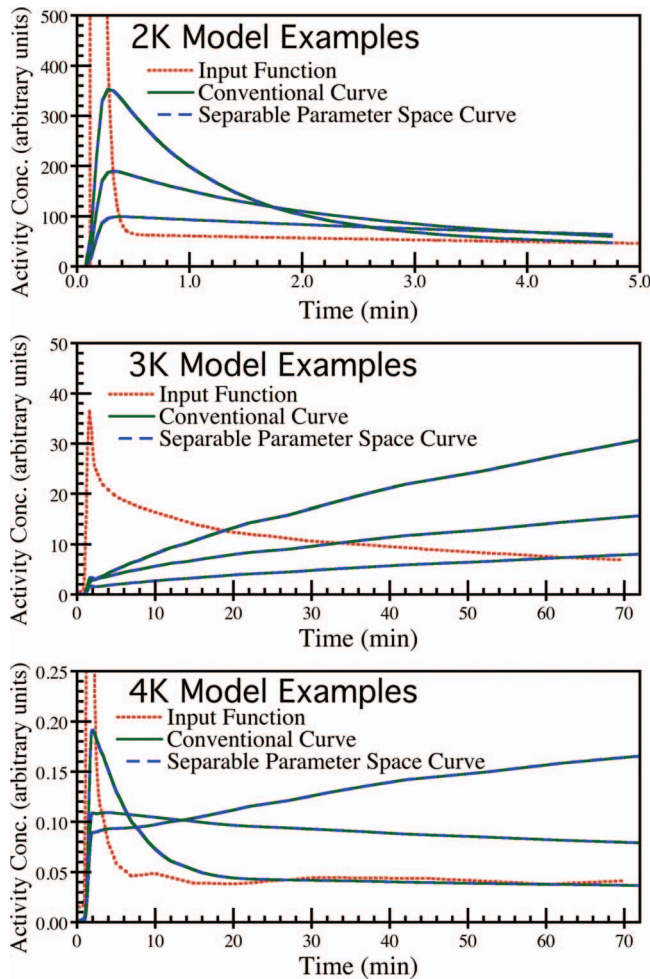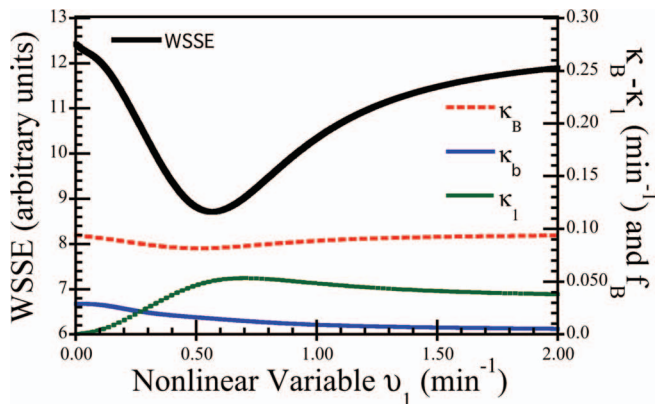
FIG. 2. Example time-activity curves for 2K-4K serial compartment models simulated using the conventional and separable parameter space model formulations. Three example curves are shown for each model. The curves for the conventional and reformulated models were identical up to the numerical precision of the computer, and overlap exactly in the plots.

IV. RESULTS

IV.A. Verification that the conventional and reformulated equations provide the same time-activity curves

Figure 2 shows example time-activity curves for 2K (based on ^{15}O-water), 3K (based on ^{18}F-FDG), and 4K (based on ^{11}C-acetate) compartment models. For each model, three representative sets of rate parameters were selected and curves were simulated using both conventional compartment model formulations and the separable parameter space formulation of Eq. (6). For the latter case, the values of κ_B, κ_b, κ_1, κ_2, υ_1, and υ_2 were computed from the conventional rate parameters according to the conversions described in Sec. II.A. The curves for each case were identical up to the numerical precision of the computer, confirming that the separable parameter space formulations correctly represent valid compartment models for these data.

IV.B. Characterization of objective functions

The separable parameter space reformulations reduce the dimensionality of the nonlinear fitting problem to only one free parameter for 2K-3K compartment models, and to only two free parameters for 4K-5K compartment models. The resultant 1D and 2D objective functions can be plotted and analyzed more easily than conventional objective functions in three to six dimensions. Figure 3 shows a representative *WSSE* objective function for a 3K compartment model fit to a noisy time-activity curve modeled on dynamic FDG PET tumor imaging. This typical objective function is well behaved, having a single well-defined minimum and no other confounding structures such as local minima or shelves. The plot also shows how the values of κ_B, κ_b, and κ_1 vary as functions of υ_1 to keep *WSSE* minimized according to the linear subproblem [Eq. (11)]; the corresponding values of f_B and the conventional rate parameters are also shown in the lower panel for comparison.
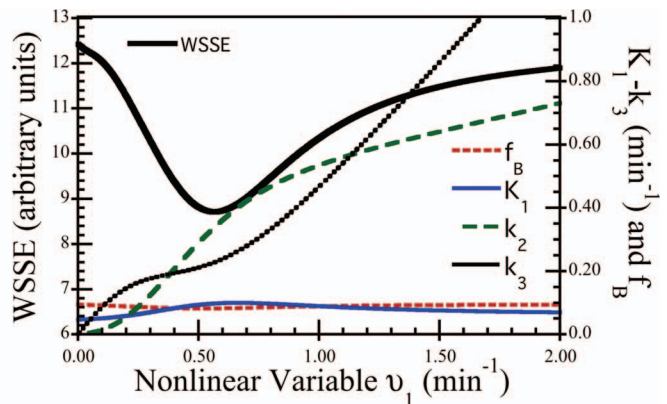

FIG. 3. Plots of the separable parameter space *WSSE* objective function for an example noisy time-activity curve with 3K compartment model. The left plot shows how the values of κ_B, κ_b, and κ_1 that minimize *WSSE* change as a function of υ_1 (recall $\upsilon_1 = k_2 + k_3$ for the 3K model). Similarly, the plot at right shows the same objective function with the corresponding values of f_B, K_1, k_2, and k_3 at each point. The objective function is well behaved, having a single clearly defined minimum with no other local minima, shelves, or confounding structures.
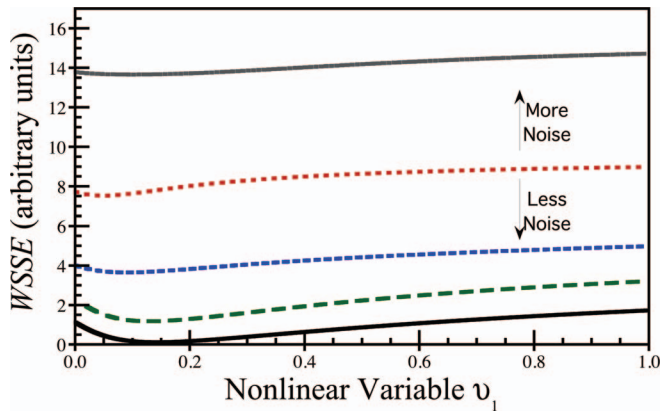
FIG. 4. Example separable parameter space *WSSE* objective functions for fitting a 3K compartment model to curves at five different noise levels. The minimum is deeper and better-defined for lower noise data, and becomes shallower as noise increases. However, for all noise levels the reformulated objective function remained well-behaved, and no complex topological features appeared with increasing noise levels.

Figure 4 shows example separable parameter space *WSSE* objective functions for another representative 3K model curve at five different noise levels. The minimum is deepest for the lowest noise curve, and becomes shallower with increasing noise—consistent with the observation that the best-fit parameters will have greater statistical variability for noisier data. However, the objective function remained well behaved for all noise levels studied. Upon reviewing hundreds of 3K model objective functions for the separable parameter space formulation, the majority of functions are well behaved with only a single minimum and little or no complicated topology; however, shallow local minima were present in a minority of cases that could affect descent-type iterative fitting algo-

rithms. In each of these cases, the local minimum occurred at relatively large values of v_1—often larger than the physiologically meaningful range—and furthermore the local minima were shallow relative to the global minimum. The potential presence of such local minima has implications for descent-type fitting algorithms, but do not affect best-fit results when the exhaustive search algorithm is used. One potential solution for handling such local minima would be to provide an upper bound constraint on the value of v_1 during the fitting procedure.

Representative objective functions in two dimensions for 4K and 5K compartment models using the separable parameter space formulation are provided in Fig. 5. Here the reformulation has two nonlinear unknowns, v_1 and v_2, both of which are constrained to be non-negative, and further the solution space was constrained to values $v_2 < v_1$ in order to limit the solution space to regions that represent valid compartment models. The majority of these 2D objective functions were again generally well-behaved, with only a single minimum, although some complexity was present for lower values of v_1 that may be considered a "side valley." These curvatures in the objective function arise from imaginary "poles" in the excluded regions $v_2 > v_1$ and $v_2 < 0$ (corresponding to values where the reformulated equations do *not* represent value compartment models—giving rise to rate parameters with either negative or imaginary values). The potential for local minima again cannot be ruled out, although empirical evidence suggests that the prevalence of local minima is low, and also that such minima are shallow and occur at relatively large values of v_1 or v_2. Overall, least-squares objective functions for the separable parameter space formulations were generally well-behaved with topology considerably less complex than that for the conventional model formulations, providing a simpler and potentially more robust environment for fast iterative nonlinear fitting.
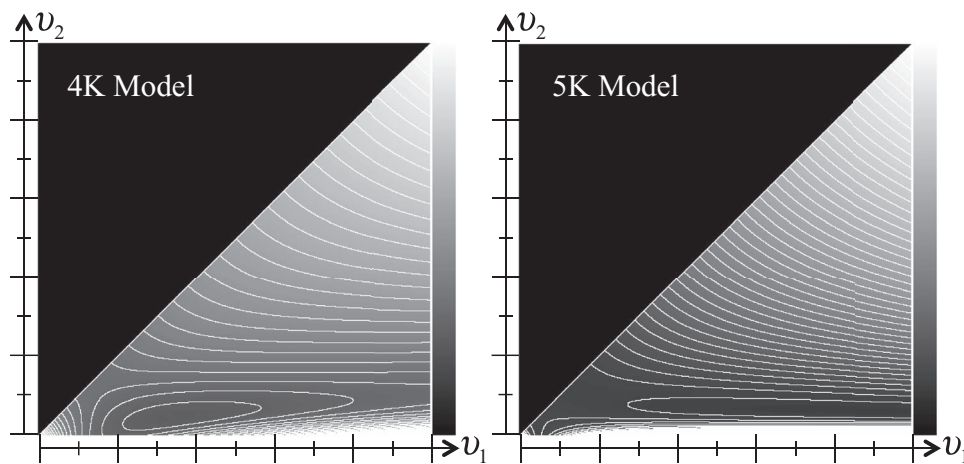

FIG. 5. Contour plots of two examples separable parameter space *WSSE* objective functions for 4K (left) and 5K (right) compartment models. The grayscale represents *WSSE* values, and contour lines are drawn at regular intervals. Recall that the separable parameter space formulation for 4K-5K models has two nonlinear unknowns (v_1, v_2), and also that v_2 is constrained to be less than v_1 (hence the plots are only defined for regions $v_2 < v_1$). The objective functions are generally well behaved and do not show complex topological features. However, some complex structure exists in the lower-left corner near $v_1 = v_2$ line and near $v_2 = 0$ which could slow convergence of gradient-descent type algorithms in these regions. These structures arise from "attraction" due to imaginary poles in the excluded $v_2 > v_1$ and $v_2 < 0$ regions of the solution space.
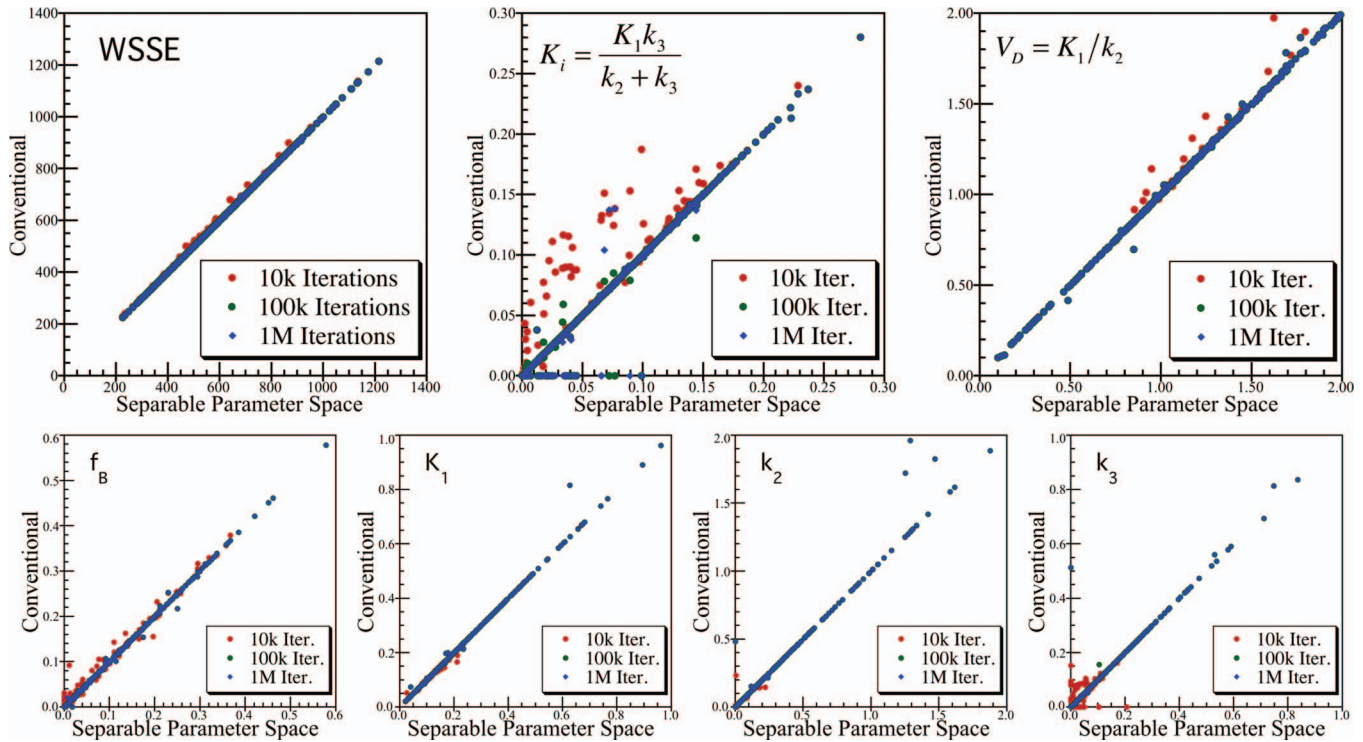
FIG. 6. Scatter plots of the 3K model *WSSE* objective function, macroparameters, and individual rate parameters comparing conventional versus separable parameter space fit results for noisy populations of time-activity curves. In each case, results for three different numbers of iterations are shown for the conventional fits with simulated annealing, demonstrating convergence effects as progressively higher numbers of iterations produced successively better fits (lower *WSSE*) for some curves. Note that in all cases the separable parameter space fits with exhaustive search produced equal or lower *WSSE* than the conventional model fits to within the numerical precision of the search (approximately 1 part in 10^{-3}). As a population, the conventional model fits approached the separable parameter space fits as the number of iterations increased.
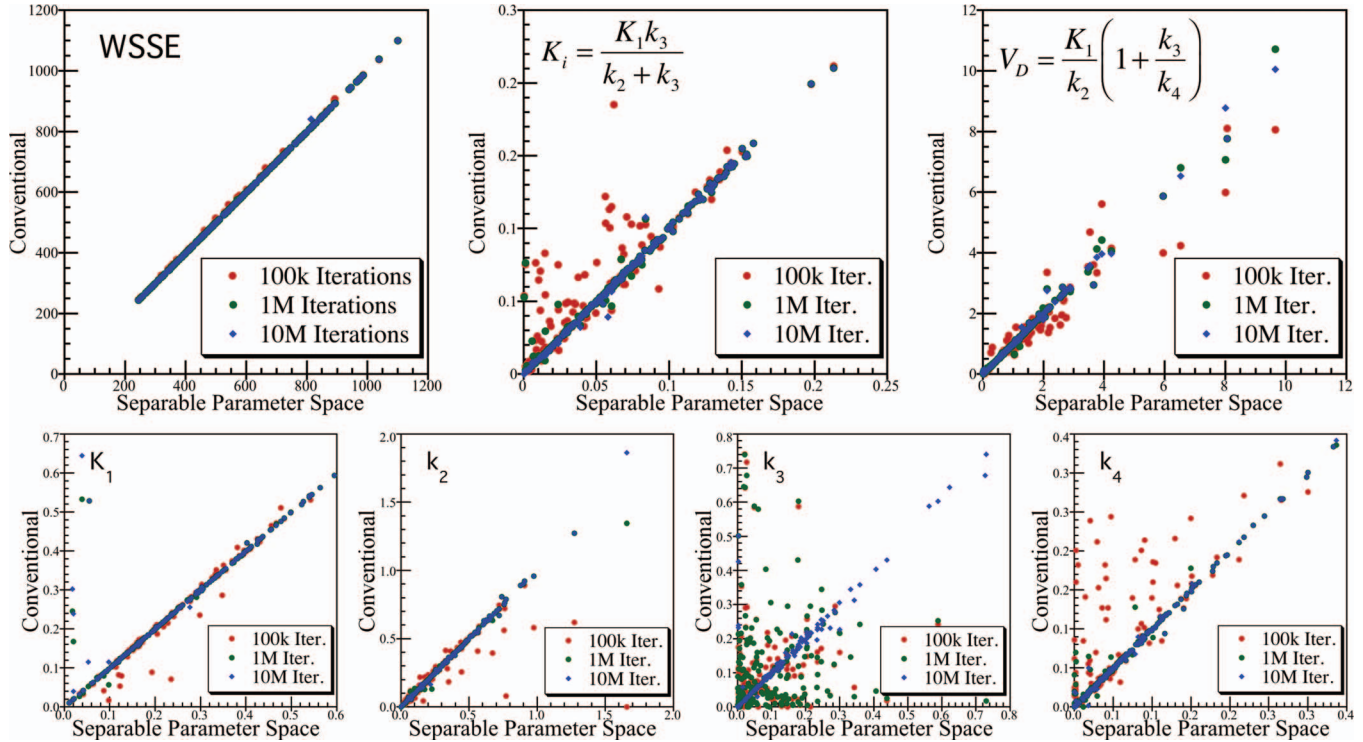

FIG. 7. Scatter plots of the 4K model *WSSE* objective function, macroparameters, and individual rate parameters (f_B not shown) comparing conventional vs separable parameter space fit results for noisy populations of time-activity curves. The conventional model results are shown for 10^5, 10^6, and 10^7 iterations of simulated annealing, showing convergence effects similar to those shown in Fig. 6 for the 3K model (but at $10\times$ more iterations due to the increased complexity of the 4K model). Again, the separable parameter space exhaustive search fits provided the lowest *WSSE* for all cases to within the precision of the search, and the conventional model fits progressive convergence toward these global minima across the full population of curves with increasing iteration.

IV.C. Verification that the conventional and reformulated model equations have the same global minimum

Populations of 200–1000 noisy time-activity curves representative of dynamic PET tumor imaging with ^{15}O-water (2K model), ^{18}F-FDG (3K model), and ^{11}C-acetate (4K model) were simulated using methods described in Sec. III.A. Each noisy curve was fit to the separable parameter space formulation using the exhaustive search algorithm with the constraints listed in Tables I and II. The curves were also fit to the conventional model formulations using simulated annealing, where the fits were repeated using 10^4–10^8 iterations progressively in 10× increments in order to progressively approach the true global minimum fits across the population of curves as the number of iterations increased.

Figure 6 shows scatter plots of the best-fit *WSSE* and rate parameters for the 3K model comparing conventional versus separable parameter space fit results. Analogous results for the 4K model are provided in Fig. 7. These figures also include results for macroparameters K_i and V_D for each model. In all cases, the separable parameter space fits with exhaustive search produced equal or lower *WSSE* than the conventional fits with simulated annealing to within the numerical precision of the exhaustive search algorithm (approximately 1 part in 10^{-3}). This result is consistent with the prediction that exhaustive search identifies the true global minimum to within the selected precision for all fits. In addition, the conventional model fits with simulated annealing progressively approached the separable parameter space exhaustive search results, further confirming that both model formulations provide the same best-fit global minimums.

Linear regression analysis between the conventional (maximum no. iterations) and separable parameter space fits revealed slopes of 1.000 ± 0.001 and Pearson correlation coefficients greater than 0.999 for all individual rate parameters and macroparameters. These results confirm that both the conventional and reformulated model equations provided identical best-fit results across these populations of curves to within the precision of the search. Small differences on the order of 10^{-4} were observed in the final rate parameters, con-

sistent with a precision on the order of 10^{-3} for the exhaustive search algorithm [note that this precision is applicable to the nonlinear parameters υ_1 and υ_2; calculation of the linear (κ) parameters and then conversion to kinetic rate parameters $K_1 - k_4$ were computed at full floating point double-precision].

IV.D. Performance of Levenberg-Marquardt algorithm

Table IV summarizes the performance of the Levenberg-Marquardt algorithm for fitting both the conventional and separable parameter space model formulations. The number of iterations required to attain the stopping criterion provide a measure of iterative convergence properties. Here, the separable parameter space fits required significantly fewer iterations to converge than did the conventional model fits, suggesting that the reduced dimensionality of the nonlinear fitting space simplified the iterative fit. In order to determine whether or not each fit converged to the global minimum, as opposed to a local minimum, the true global minimum fit for each case was identified from the simulated annealing and exhaustive search results from Sec. IV.C. For each Levenberg-Marquardt fit, the difference in best-fit *WSSE* between the Levenberg-Marquardt fit and the global minimum was computed. The median value and range are shown in the table. Similarly, the best-fit kinetic parameters (f_B, $K_1 - k_4$) for the Levenberg-Marquardt fits were evaluated versus those for the global minimum fits by computing the magnitude of the error vector (root sum-square error in parameter estimates). Again, the median value and range of the error vector magnitude are listed in the table.

The fit results were sensitive to the initial conditions for both the conventional and separable parameter space model formulations, indicating that local minima may be present (at least some of the time) for both cases. However, the separable parameter space fits were less sensitive to initial conditions than were the conventional model fits. In particular, initializing the fit with midrange parameter values provided excellent results. When using these initial values, the separable parameter space fits attained the true global minimum fit for all

TABLE IV. Levenberg-Marquardt fit results.

		Conventional model fits			Separable parameter space fits		
Model	Initial condition	Number of iterations (mean ± SD)	WSSE difference (median [range])	Magnitude error vector (median [range])	Number of iterations (mean ± SD)	WSSE difference (median [range])	Magnitude error vector (median [range])
2K	Random	314 ± 390	314 [0–76705]	5.300 [0–6.97]	19 ± 16	0.00 [0–76706]	0.001 [0.00–4.97]
	Minimum	205 ± 349	205 [0–814]	1.900 [0–2.88]	20 ± 24	0.00 [0–18.9]	0.001 [0.00–1.99]
	Midrange	123 ± 250	167 [0–1045]	1.100 [0–1.32]	20 ± 6	0.00 [0–0]	0.000 [0.00–0.11]
3K	Random	117 ± 260	3214 [0–136933]	2.990 [0.0–6.05]	15 ± 14	26 [0–695]	0.084 [0.00–7.14]
	Minimum	906 ± 234	107 [0–1393]	1.909 [0.062–4.82]	15 ± 18	32 [0–285]	2.881 [0.00–4.38]
	Midrange	117 ± 178	155 [10–2566]	1.008 [0.065–3.56]	28 ± 18	0.00 [0.0–0.0]	0.000 [0.00–0.002]
4K	Random	287 ± 389	1593 [0–56568]	2.357 [0.01–5.39]	130 ± 177	282 [0–16545]	1.550 [0.00–3.23]
	Minimum	723 ± 422	460 [0–54330]	1.894 [0.02–2.94]	62 ± 171	0.00 [0–796]	0.004 [0.00–2.79]
	Midrange	483 ± 460	16 [0–245]	1.478 [0.00–2.01]	80 ± 124	0.00 [0.0–0.0]	0.000 [0.00–0.01]

TABLE V. CPU times for fits.

Model	2K		3K		4K	
	Number of iterations	CPU time (ms)	Number of iterations	CPU time (ms)	Number of iterations	CPU time (ms)
Conventional model formulations						
Simulated annealing	10^4	326 ± 22	10^5	2227 ± 163	10^6	24056 ± 4508
Levenberg-Marquardt	123 ± 250	6.87 ± 13.97	117 ± 178	8.79 ± 13.39	483 ± 460	48.80 ± 46.19
Separable parameter space reformulations						
Exhaustive search	10^3	10.2 ± 1.5	10^3	10.1 ± 0.9	10^6	1109 ± 172
Levenberg-Marquardt	20 ± 6	0.90 ± 0.27	28 ± 18	1.32 ± 0.85	80 ± 124	7.34 ± 11.38

parameters (defined as magnitude of the best-fit error vector < 0.01) in 96% of cases for the 2K model, 98% of cases for the 3K model, and 94% of cases for the 4K model. In contrast, the conventional model fits attained the true global minimum fit for all parameters in 53% of cases for the 2K model, in 4.8% of cases for the 3K model, and in 1.5% of cases for the 4K model. These results demonstrate the importance of manually selecting good initial conditions for the Levenberg-Marquardt algorithm with the conventional model formulation. Conversely, the Levenberg-Marquardt fits for the separable parameter space formulation were much less sensitive to initial conditions, which can be seeded within midrange initial values and converge to the true global minimum in the majority of cases.

IV.E. Fitting times

Table V shows the CPU times for fitting noisy time-activity curves with 2K, 3K, and 4K models using both the conventional and separable parameter space model formulations. Times were computed for single-threaded code run on a 2.80 GHz Intel Xeon workstation, and no acceleration via multithreading was implemented. Each algorithm was implemented in C, sharing routines when possible, and with identical attention to optimizing the code for computational efficiency. The CPU time per iteration was similar for both model formulations, on the order of 0.04–0.07 ms for 2K-3K models and 0.09–0.10 ms for 4K models. Overall, the separable parameter space fits required approximately 10 ms for 2K-3K models and 1.1 s for 4K models; these times were reduced to about 1 and 7 ms, respectively, using the Levenberg-Marquardt algorithm.

V. DISCUSSION AND CONCLUSIONS

The separable parameter space techniques discussed in this paper generalize previously proposed methods for applying separable nonlinear least-squares to the compartment model fitting problem. The formulation described herein encompasses 1K-5K serial compartment models, including a modified formulation for higher order models that avoids a potential degeneracy present in previously proposed reformulations for these models. As a theoretical exercise, the linear subproblem of the reformulated equations was solved analytically and substituted back into the remaining nonlinear fitting equations. The resultant system can be interpreted as effectively reducing the dimensionality of the fitting space to only one dimension for 2K-3K compartment models, and to two dimensions for 4K-5K models—providing a nonlinear fitting problem that is both simpler and easier to solve than that of the original model formulation. This illustrates the value of the separable parameter space technique, but is not amenable to imposing constraints on the linear parameters. When keeping the linear subproblem separate, a fast and robust algorithm was presented for solving the linear subproblem with complete user-defined constraints on the fitted parameters. The generalized technique also treats the nonlinear unknowns as fittable parameters, rather than treating their corresponding temporal terms as basis functions. The receptiveness of the technique to fast gradient-descent type nonlinear fitting algorithms was explored by characterizing the objective functions, and by studying the performance of the Levenberg-Marquardt algorithm for solving the nonlinear subproblem.

The work focused specifically on the topic of performing the mathematical fit according to a preestablished fitting criterion—the WLS criterion. In order to retain focus upon the mathematical fitting approach, we did not explore other issues such as consideration of other fitting criteria, selection of the fitting weights, or identification of the input function. While these issues affect the bias and noise properties of the fitted parameters, they are distinct from issues regarding the mathematical fit—the fitting algorithm is a mathematical procedure for attaining the fitting solution that minimizes the fitting criterion, whereas the criterion, weights, and input function determine the properties of the fit at the global minimum.

When using the Levenberg-Marquardt algorithm, it is a common practice to estimate the covariance matrix of the standard errors in the fitted parameters by inverting the Hessian matrix used by the fitting algorithm[33] (typically with second derivatives ignored). When using separable parameter space techniques, the Hessian matrix used in the nonlinear subproblem corresponds to only the nonlinear free parameters [as well as being subject to the constraint of minimizing the linear subproblem, as per Eq. (12)]. In order to estimate the covariance matrix of the fitted kinetic rate parameters, one should compute and invert the Hessian matrix of the

conventional full-parameter space model formulation using the best-fit rate parameters obtained from the separable parameter space fit. This approach is viable since both the conventional and separable parameter space formulations have the same global minimum, and requires a small degree of additional computation after completing the fit.

The generalized separable parameter space technique was found to provide identical time-activity curves as conventional compartment model formulations, and both the conventional and reformulated equations give rise to the same global minima. As such, fitting the reformulated equations (with appropriate constraints) and calculating conventional rate parameters from the best-fit reformulated parameters effectively finds the best-fit to the conventional model equations. However, the reduced dimensionality of the reformulated nonlinear fitting subproblem greatly simplifies separable parameter space fits. This reduced dimensionality makes exhaustive search algorithms feasible (and indeed computationally rapid), providing fits that guarantee the true global minimum to within the search precision in as little 10 ms for 2K-3K models and 1.1 s for 4K models. For applications in which more rapid fitting is desired, e.g., voxelwise parametric imaging, fast iterative fitting algorithms, such as Levenberg-Marquardt, can be used to reduced the fitting times to approximately 1 ms for 2K-3K models and 7 ms for 4K models. Such fits were found to be less sensitive to initial conditions than comparable iterative fits with the conventional model formulations, and to converge in fewer iterations. In conclusion, the separable parameter space techniques described in this work generalize previously described techniques to encompass 1K-5K compartment models, enable robust solution of the linear subproblem with full user-defined constraints, and are amenable to rapid and robust fitting using iterative gradient-descent type algorithms.

ACKNOWLEDGMENTS

APPENDIX: PSEUDOCODE ALGORITHM FOR COMPUTING CONSTRAINED LINEAR SUBPROBLEM

Let $\kappa_i^{\min}$ and $\kappa_i^{\max}$ be arbitrary user-defined minimum and maximum values for each linear parameter κ_i.

The linear subproblem described in Sec. II.C can be rapidly solved subject to these constraints using the following pseudocode algorithm:

(1) Analytically invert the linear subproblem directly to obtain unconstrained best-fit values of all κs.

(2) If any κ_i is out of range, where $i \in \{B, 2, 1, b\}$, then:

Loop over $\kappa_i^{\min} \le \kappa_i \le \kappa_i^{\max}$ with step size Δ_i: (i.e., exhaustively search over κ_i; typical $\Delta_i = 0.001$ min^{-1})

Fixing κ_i, analytically invert the remaining linear subproblem for the remaining free κs.

If any κ_j is out of range, where $j \in \{B, 2, 1, b\}$ and $j \ne i$, then:

Loop over $\kappa_j^{\min} \le \kappa_j \le \kappa_j^{\max}$ with step size Δ_j: (i.e., exhaustively search over κ_j)

Fixing κ_j, analytically invert the remaining linear subproblem for the remaining free κs.

If any κ_k is out of range, where $k \in \{B, 2, 1, b\}$ and $k \ne i$ and $k \ne j$, then:

Loop over $\kappa_k^{\min} \le \kappa_k \le \kappa_k^{\max}$ with step size Δ_k: (i.e., exhaustively search over κ_k)

Analytically compute κ_l, where $l \in \{B, 2, 1, b\}$ and $l \ne i$, $l \ne j$, and $l \ne k$.

If κ_l is out of range, then:

Loop over $\kappa_l^{\min} \le \kappa_l \le \kappa_l^{\max}$ with step size Δ_l: (i.e., exhaustively search over κ_l)

Find κ_l that maximizes *WSSE* for the current values of κ_i, κ_j, and κ_k.

End loop.

End if.

Find κ_k that maximizes *WSSE* for the current values of κ_i, κ_j, and κ_l.

End loop.

End if.

Find κ_j that maximizes *WSSE* for the current values of κ_i, κ_k, and κ_l.

End loop.

End if.

Find κ_i that maximizes *WSSE* for the current values of κ_j, κ_k, and κ_l.

End loop.

End if.

[a] Author to whom correspondence should be addressed. Electronic mail: kadrmas@ucair.med.utah.edu

[1] R. E. Carson, "Tracer kinetic modeling in PET," in *Positron Emission Tomography*, edited by D. L. Bailey, D. W. Townsend, P. E. Valk, and M. N. Maisey (Springer-Verlag, London, 2005).

[2] S. R. Cherry, J. A. Sorenson, and M. E. Phelps, *Physics in Nuclear Medicine* (Saunders, Philadelphia, 2003).

[3] K. C. Schmidt and F. E. Turkheimer, "Kinetic modeling in positron emission tomography," Q. J. Nucl. Med. **46**, 70–85 (2002).

[4] H. Watabe, Y. Ikoma, Y. Kimura, M. Naganawa, and M. Shidahara, "PET kinetic analysis–Compartmental model," Ann. Nucl. Med. **20**, 583–588 (2006).

[5] S. Urien, "MicroPharm-K: A microcomputer interactive program for the analysis and simulation of pharmacokinetic processes," Pharm. Res. **12**, 1225–1230 (1995).

[6] X. Zhang, J. N. Andrews, and S. E. Pedersen, "Kinetic modeling and fitting software for interconnected reaction schemes: VisKin," Anal. Biochem. **361**, 153–161 (2007).

[7] F. O'Sullivan and A. Saha, "Use of ridge regression for improved estimation of kinetic constants from PET data," IEEE Trans. Med. Imaging **18**, 115–125 (1999).

[8] Y. Zhou, H. Sung-Cheng, and M. Bergsneider, "Linear ridge regression with spatial constraint for generation of parametric images in dynamic positron emission tomography studies," IEEE Trans. Nucl. Sci. **48**, 125–130 (2001).

[9] Y. Zhou, C. J. Endres, J. R. Brasic, S. C. Huang, and D. F. Wong, "Linear regression with spatial constraint to generate parametric images of ligand-receptor dynamic PET studies with a simplified reference tissue model," Neuroimage **18**, 975–989 (2003).

[10] M. Byrtek, F. O'Sullivan, M. Muzi, and A. M. Spence, "An adaptation of ridge regression for improved estimation of kinetic model parameters from PET studies," IEEE Trans. Nucl. Sci. **52**, 63–68 (2005).

[11] X. Dai, Z. Chen, and J. Tian, "Performance evaluation of kinetic parameter estimation methods in dynamic FDG-PET studies," Nucl. Med. Commun. **32**, 4–16 (2011).

[12] A. Eftaxias, J. Font, A. Fortuny, A. Fabregat, and F. Stuber, "Nonlinear kinetic parameter estimation using simulated annealing," Comput. Chem. Eng. **26**, 1725–1733 (2002).

[13] K. P. Wong, S. R. Meikle, D. Feng, and M. J. Fulham, "Estimation of input function and kinetic parameters using simulated annealing: Application in a flow model," IEEE Trans. Nucl. Sci. **49**, 707–713 (2002).

[14] M. Yaqub, R. Boellaard, M. A. Kropholler, and A. A. Lammertsma, "Optimization algorithms and weighting factors for analysis of dynamic PET studies," Phys. Med. Biol. **51**, 4217–4232 (2006).

[15] R. E. Marsh, T. A. Riauka, and S. A. McQuarrie, "Use of a simulated annealing algorithm to fit compartmental models with an application to fractal pharmacokinetics," J. Pharm. Pharm. Sci. **10**, 168–179 (2007).

[16] W. H. Lawton and E. A. Sylvestre, "Elimination of linear parameters in nonlinear regression," Technometrics **13**, 461–467 (1971).

[17] L. Kaufman and V. Pereyra, "A method for separable nonlinear least squares problems with separable nonlinear equality," SIAM J. Numer. Anal. **15**, 12–20 (1978).

[18] G. H. Golub and V. Pereyra, "The differentiation of pseudo-inverses and nonlinear least squares problems whose variables separate," SIAM J. Numer. Anal. **10**, 413–432 (1973).

[19] G. Golub and V. Pereyra, "Separable nonlinear least squares: The variable projection method and its applications," Inverse Probl. **19**, R1–R26 (2003).

[20] R. A. Koeppe, J. E. Holden, and W. R. Ip, "Performance comparison of parameter estimation techniques for the quantitation of local cerebral blood flow by dynamic positron computed tomography," J. Cereb. Blood Flow Metab. **5**, 224–234 (1985).

[21] R. H. Huesman, B. W. Reutter, G. L. Zeng, and G. T. Gullberg, "Kinetic parameter estimation from SPECT cone-beam projection measurements," Phys. Med. Biol. **43**, 973–982 (1998).

[22] B. W. Reutter, G. T. Gullberg, and R. H. Huesman, "Kinetic parameter estimation from attenuated SPECT projection measurements," IEEE Trans. Nucl. Sci. **45**, 3007–3013 (1998).

[23] R. Boellaard, P. Knaapen, A. Rijbroek, G. J. Luurtsema, and A. A. Lammertsma, "Evaluation of basis function and linear least squares methods for generating parametric blood flow images using 15O-water and positron emission tomography," Mol. Imaging Biol. **7**, 273–285 (2005).

[24] H. Watabe, H. Jino, N. Kawachi, N. Teramoto, T. Hayasi, Y. Ohta, and H. Iida, "Parametric imaging of myocardial blood flow with 15O-water and PET using the basis function method," J. Nucl. Med. **46**, 1219–1224 (2005).

[25] R. N. Gunn, A. A. Lammertsma, S. P. Hume, and V. J. Cunningham, "Parametric imaging of ligand-receptor binding in PET using a simplified reference region model," Neuroimage **6**, 279–287 (1997).

[26] R. N. Gunn, S. R. Gunn, F. E. Turkheimer, J. A. Aston, and V. J. Cunningham, "Positron emission tomography compartmental models: A basis pursuit strategy for kinetic modeling," J. Cereb. Blood Flow Metab. **22**, 1425–1439 (2002).

[27] Y. T. Hong and T. D. Fryer, "Kinetic modelling using basis functions derived from two-tissue compartmental models with a plasma input function: General principle and application to [18F]fluorodeoxyglucose positron emission tomography," Neuroimage **51**, 164–172 (2010).

[28] Y. T. Hong, J. S. Beech, R. Smith, J. C. Baron, and T. D. Fryer, "Parametric mapping of [18F]fluoromisonidazole positron emission tomography using basis functions," J. Cereb. Blood Flow Metab. **31**, 648–657 (2011).

[29] D. Feng, D. Ho, K. Chen, L. C. Wu, J. K. Wang, R. S. Liu, and S. H. Yeh, "An evaluation of the algorithms for determining local cerebral metabolic rates of glucose using positron emission tomography dynamic data," IEEE Trans. Med. Imaging **14**, 697–710 (1995).

[30] M. E. Phelps, *PET: Molecular Imaging and Its Biological Applications* (Springer Science, New York, NY, 2004).

[31] A. Bertoldo, P. Peltoniemi, V. Oikonen, J. Knuuti, P. Nuutila, and C. Cobelli, "Kinetic modeling of [(18)F]FDG in skeletal muscle by PET: A four-compartment five-rate-constant model," Am. J. Physiol. Endocrinol. Metab. **281**, E524–E536 (2001).

[32] G. Tomasi, A. Bertoldo, S. Bishu, A. Unterman, C. B. Smith, and K. C. Schmidt, "Voxel-based estimation of kinetic model parameters of the L-[1-(11)C]leucine PET method for determination of regional rates of cerebral protein synthesis: Validation and comparison with region-of-interest-based methods," J. Cereb. Blood Flow Metab. **29**, 1317–1331 (2009).

[33] W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes in C* (Cambridge University Press, Cambridge, 1988).