Annals of Biomedical Engineering, Vol. 37, No. 5, May 2009 (© 2009) pp. 1028–1042
DOI: 10.1007/s10439-009-9651-z

Parameter Estimation for Linear Compartmental Models—A Sensitivity Analysis Approach

BARBARA JUILLET, CÉCILE BOS, CLAIRE GAUDICHON, DANIEL TOMÉ, and HÉLÈNE FOUILLET

UMR914 Nutrition Physiology and Ingestive Behavior, INRA, AgroParisTech, CRNH-IdF, 16 rue Claude Bernard, F-75005 Paris, France

(*Received 4 December 2007; accepted 6 February 2009; published online 19 February 2009*)

Abstract—Linear compartmental models are useful, explanatory tools, that have been widely used to represent the dynamic behavior of complex biological systems. This paper addresses the problem of the numerical identification of such models, i.e., the estimation of the parameter values that will generate predictions closest to experimental observations. Traditional local optimization techniques find it difficult to arrive at satisfactory solutions to such a parameter estimation problem, especially when the number of parameters is large and/or few data are available from experiments. We present herewith a method based on a prior sensitivity analysis, which enables division of a large optimization problem into several smaller and simpler subproblems, on which only sensitive parameters are estimated, before the whole optimization problem is tackled from starting points that are already close to the optimum values. This method has been applied successfully to a linear 13-compartment, 21-parameter model describing the postprandial metabolism of dietary nitrogen in humans. The effectiveness of the method has been demonstrated using simulated and real data obtained in the intestine, blood and urine of healthy humans after the ingestion of a [^{15}N]-labeled protein meal.

Keywords—Biological system modeling, Inverse problem, Parameter estimation, Optimization methods, Sensitivity, Algorithms, Compartmental models, Nitrogen metabolism, Dietary proteins, Tissue distribution.

ABBREVIATIONS

BU	body urea
CV	coefficient of variation
d_0	noise-free data
d_1	low homogeneous noise data
d_2	high homogeneous noise data
d_3	high heterogeneous noise data
E	ileal effluents
G	gastric content
IL_1	proximal intestinal lumen
IL_2	distal intestinal lumen
LLF	log of the likelihood function
N	nitrogen
ODE	ordinary differential equations
PAA	peripheral free amino acids
PP	peripheral proteins
SA	sensitivity analysis
SAA	splanchnic free amino acids
SCP	splanchnic constitutive proteins
SEP	splanchnic exported proteins
UU	urinary urea
UA	urinary ammonia

INTRODUCTION

It has now become standard practice to use multicompartmental models to study complex dynamic systems in many research fields of the biological sciences. In particular, compartmental models have been widely employed to investigate the distribution of materials in living systems, providing valuable insight into the metabolism of numerous nutrients in humans, including micronutrients,[27] glucose,[38] lipids,[46] amino acids[12] and proteins.[20] In the research area of human protein metabolism, linear compartmental models have regularly been proposed in recent decades to describe the metabolism of a single amino acid,[4,8,12,44] or total dietary nitrogen (N) under various nutritional and/or physiopathological conditions.[17–20] These models have enabled a clearer understanding of the different metabolic processes involved in protein and amino acid homeostasis in humans, such as muscle protein synthesis and breakdown,[4,44] the contribution of splanchnic protein pools to total postprandial anabolism,[8,18] and the nutritional modulation of the regional metabolism of dietary N.[17,19]

Address correspondence to Hélène Fouillet, UMR914 Nutrition Physiology and Ingestive Behavior, INRA, AgroParisTech, CRNH-IdF, 16 rue Claude Bernard, F-75005 Paris, France. Electronic mail: Helene.Fouillet@agroparistech.fr

Based on mass conservation principles, linear compartmental models are very attractive to users because knowledge and intuition are formalized in a simple, reasonable fashion: compartments represent the amount of the studied compound present at different sites and/or in different forms in the body, and they exchange material through various transfer and/or metabolic pathways described using simple, linear diffusion laws.[24] Furthermore, this modeling approach makes it possible to reduce complex physiological systems into a finite number of compartments and pathways, thus restricting mismatches between the complexity of the system and the few data available from *in vivo* studies, especially in humans.[10] Development of such models involves two principal steps: (i) determination of their structure, i.e., the number and localization of the model compartments and exchange pathways that will best represent the system under study, and (ii) numerical estimation of their parameter values by fitting model predictions to experimental observations of the system.[24] Given a model structure, step (ii) of its numerical identification still constitutes a challenging mathematical problem whose difficulty rapidly increases with model complexity.[37]

In compartmental modeling, the methods used to carry out parameter estimation have markedly improved over the past thirty years, in parallel with the development of powerful tools for numerical calculation. Initially, model parameters were manually and approximately adjusted to achieve a visually acceptable match between model predictions and experimental data. Nowadays, modeling softwares incorporate automated and more reliable parameter estimation techniques: an objective function measuring the quality of model predictions (usually the least squares or likelihood function) is optimized to determine parameter values that will generate predictions closest to experimental observations.[11,41] This optimization process is generally conducted using a local, gradient-based algorithm: starting from user-specified parameter values, it iteratively attempts to locate an optimum point in the surface defined by the objective function. Nevertheless, this surface becomes extremely complicated when the number of model parameters is large and/or few data are available from experiments.[20,41] In such cases, the search algorithm usually converges to local optima, especially when the starting values of certain adjustable parameters are too distant from the correct ones.[40] As a consequence, a lack of knowledge on the studied metabolic or nutritional system often results in poor initial guesses for model parameters (i.e., outside the basin of attraction of the global optimum) and mostly leads to a failure of the optimization algorithm.[37]

The simplest method generally used to overcome this limitation is based on the idea of starting the local algorithm repeatedly from different initial sets of parameter values in order to explore the parameter space intensively.[22] However, this method is known to be inefficient in practice: when several initial points are used, the same local minima may be found several times.[22,31] In such a context, global optimization methods have often been presented as an alternative, even if there are weak theoretical guarantees of convergence to the global optimum.[29] In particular, evolutionary algorithms such as genetic algorithms, evolutionary programming and evolution strategies, and other global optimization methods including clustering methods, simulated annealing and tabu search have recently been tested to solve a large inverse problem regarding a biochemical kinetic system.[31] Only a certain type of these stochastic global optimization methods (evolution strategies) was able to solve the inverse problem successfully, although at a large computational cost. The authors have thus recently turned to hybrid global-local methods to reduce the computational time of their estimation method while maintaining its robustness.[37] Another, and maybe more intuitive, strategy has also been suggested; this consists in using information drawn from a sensitivity analysis (SA) of the model to simplify the parameter estimation process.[20,41] In these studies, the prior identification of sensitive parameters (i.e., parameters with the greatest influence on model outputs) enabled a focus on their variations during optimization: the sensitive parameters are estimated in the first instance and the effect of their conjoint variations on model fitting is explored extensively before estimating all parameters simultaneously. As other local or global optimization techniques, this method does not guarantee convergence to the global optimum but offers the advantage of providing valuable information about system behavior and dynamics. Although such a prior SA has already been claimed to facilitate the process of parameter estimation, little or no information has generally been provided concerning its practical implementation and how it makes the parameter estimation process easier. Indeed, the use of information drawn from SA is rarely fully transparent in the literature. There is a lack of—and thus a need for—a formal description of how to use this sensitivity information in order to identify compartmental models.

In the present work, we propose a rationalization and generalization of this SA approach, so that it can be applied routinely to estimating the parameter values for any linear compartmental model. The method is based on a prior SA which will identify those parameters which can (or cannot) drive significant variations

to the different system outputs. The sensitivity results are then used to facilitate parameter estimation: the whole optimization problem is divided into several simpler, nested subproblems, on which only sensitive parameters are estimated, before searching for all parameter values from starting points that are already close to their optimum values. This method has been applied successfully to a linear 13-compartment, 21-parameter model describing dietary N metabolism in humans,[18] using first simulated data and then real clinical data obtained in the intestine, blood and urine of healthy humans after the ingestion of a [^{15}N]-labeled protein meal.

SYSTEM MODEL AND INVERSE PROBLEM

Forward Problem

A linear compartmental model, together with the input-output experiment designed for its identification, is generally described in a state-space form, as follows[7,24]:

$$\dot{\mathbf{x}}(t,\mathbf{p}) = A(\mathbf{p})\mathbf{x}(t,\mathbf{p}) + B(\mathbf{p})\mathbf{u}(t), \tag{1}$$

$$\mathbf{y}(t,\mathbf{p}) = C(\mathbf{p})\mathbf{x}(t,\mathbf{p}), \tag{2}$$

where $\mathbf{x}$ is the n-dimensional vector of state variables (model compartment sizes), $\mathbf{u}$ the r-dimensional input vector (e.g., intravenous injections, oral load, etc.), $\mathbf{y}$ the m-dimensional output vector (sampled compartment sizes), A the $n \times n$ compartmental matrix containing the constant transfer rates of the model, B an $n \times r$ input matrix, C an $m \times n$ output matrix and $\mathbf{p} \in \mathcal{P}$ a v-dimensional vector containing the A, B and C unknown parameters. If initial conditions are specified, the relevant equation $\mathbf{x}(t_0) = \mathbf{x}_0$ is added to the system. If equality constraints (linear or nonlinear) on $\mathbf{p}$ exist, they are embedded directly into the model.

Equation (1) is a first-order linear ordinary differential equation (ODE), which is mostly too complex to be solved analytically and has to be solved numerically using an ODE solver. Some ODE solvers (Gear, Rosenbrock, etc.) are specifically designed for stiff problems, i.e., for problems that have both small and large time constants, while classic integrators (Runge-Kutta, Adams-Bashforth-Moulton, etc.) may be inefficient. Solving Eq. (1) enables evaluation of all model state variables at each time step over the time period indicated. Equation (2) describes the temporal evolution of model outputs, i.e., the state variables that are observed experimentally.

In practice, only a small number of state variables can be monitored experimentally when studying a complex nutritional or metabolic system, since only a reduced number of metabolic pools are generally accessible to experimental measurement, especially in humans.[17] The problem facing the biologist is then to extract as much information as possible from these scarce and noisy data to determine the model structure and its parameter values, i.e., the values that will enable the model to reproduce satisfactorily the measured responses.[7,24] In the present work, we will focus on the parameter estimation problem (also known as the inverse problem), assuming the model structure as given.

Inverse Problem

Assume that experimental measurements are taken in the presence of zero mean Gaussian noise $\mathbf{e}(t)$, uncorrelated in time and space and with variance σ^2. The measurement model is then given by[11]:

$$\mathbf{z}(t) = \mathbf{y}(t,\mathbf{p}) + \mathbf{e}(t),$$

where $\mathbf{z}(t)$ denotes the m-dimensional measurement vector obtained from experiments.

In compartmental modeling, the objective function ϕ quantifying the quality of model responses is usually defined as a measure of the differences between model predictions and experimental observations[11]:

$$\phi(\mathbf{p}) = \sum_{j=1}^{m}\sum_{i=1}^{n_j} \omega(\epsilon_{ij},\mathbf{p}),$$

where m is the number of sampled compartments, n_j the number of sampling times for compartment j, ω an arbitrary cost function (usually the least squares or logarithm of the likelihood function) and ϵ_{ij} an appropriately weighted residual:

$$\epsilon_{ij} = \frac{\mathbf{e}_j(t_i)}{\sigma_{ij}},$$

where $\mathbf{e}_j(t_i)$ is the measurement error made on compartment j at time t_i and σ_{ij} its standard deviation.

In some cases, measurement errors have an experimentally determined variance.[38] However, the experimental deduction of variances in error measurements is often impossible, e.g., when the experiment cannot be reproduced numerous times in the same subject to obtain several replicates of the same sample. In such cases, a mathematical model can be postulated to represent the variability of experimental measurements, for instance[20,24]:

$$\sigma_{ij}^2 = \omega_j^2 \mathbf{y}_j^{\gamma_j}(t_i,\mathbf{p}), \tag{3}$$

with ω_j a proportionality factor and γ_j a heteroscedasticity parameter. The degree of dependence of σ_{ij} on the predicted value $\mathbf{y}_j(t_i,\mathbf{p})$ is determined by the value

of γ_j between 0 and 2: if $\gamma_j = 0$, σ_{ij} is independent of $\mathbf{y}_j(t_i, \mathbf{p})$, whereas σ_{ij} is directly proportional to $\mathbf{y}_j(t_i, \mathbf{p})$ if $\gamma_j = 2$. Coefficients ω_j and γ_j then become additional parameters to be estimated during the fitting process.

Solving this problem requires optimization of the objective function ϕ, which is classically achieved by either minimizing the residuals between measured and simulated data (least squares method), or maximizing the joint probability of having obtained the experimental data for all sampled compartments simultaneously (maximum likelihood method). When standard methods of local optimization are used, the most common pitfall is convergence to local solutions, especially when the number of parameters is large and/or when only poor starting values are available.[24,40] However, it is generally of particular importance to determine the true parameter values of a physiological model, because (i) they are informative in themselves (values of the rates of absorption, transfer or metabolic use of the studied compound) and (ii) different sets of parameter values may give rise to different physiological conclusions and interpretations.[9,16] There is thus an obvious need for developing optimization methods that facilitate convergence to the globally optimal solution.

PARAMETER ESTIMATION METHOD

We present herewith a method which divides a large and hardly solvable optimization problem into smaller and more easily solvable subproblems. This procedure uses a prior SA to identify the most relevant subproblems which need to be solved.

Prior SA

SA studies how the variation in the outputs of a model can be apportioned to different sources of variation, for example model parameters, also called input variables. Different approaches may be used to quantify the sensitivities of model outputs to changes in input parameters. In the present work, a variance-based global SA is suggested,[42] that can be decomposed in the following steps:

(1) Definition of the possible range of variation of each model parameter;
(2) Generation of a large sample of parameter values that are uniformly distributed over the parameter space;
(3) Evaluation of the model outputs for each element of the sample;
(4) Estimation of the effect of each input variable on the model outputs.

The last step is performed using a decomposition of the output variance into a finite number of factorial terms of increasing dimensionality[36,39,42]:

$$V(\mathbf{y}_j(t_i)) = \sum_{k=1}^{s} D^k + \sum_{1 \le k < l \le s} D^{kl} + \cdots + D^{12\ldots s},$$

where s is the number of model parameters, $V(\mathbf{y}_j(t_i))$ is the total variance of the output variable $\mathbf{y}_j$ at time t_i, $D^k = V(E(\mathbf{y}_j(t_i)|\mathbf{p}_k))$ is the variance over all values of $\mathbf{p}_k$ in the conditional expectation of $\mathbf{y}_j(t_i)$ given $\mathbf{p}_k$, $D^{kl} = V(E(\mathbf{y}_j(t_i)|\mathbf{p}_k, \mathbf{p}_l)) - V(E(\mathbf{y}_j(t_i)|\mathbf{p}_k)) - V(E(\mathbf{y}_j(t_i)|\mathbf{p}_l))$, and so on. The sensitivity indices are computed as follows:

$$S_j^k(t_i) = \frac{D^k}{V(\mathbf{y}_j(t_i))}, \tag{4}$$

$$S_j^{kl}(t_i) = \frac{D^{kl}}{V(\mathbf{y}_j(t_i))}, \tag{5}$$

$$S_j^{1,\ldots,s}(t_i) = \frac{D^{12\ldots s}}{V(\mathbf{y}_j(t_i))}, \tag{6}$$

where $S_j^k(t_i)$ is the fractional contribution of $\mathbf{p}_k$ to the total variance of the output variable $\mathbf{y}_j$ at time t_i (also called importance measure or first-order sensitivity index), $S_j^{kl}(t_i)$ is the fractional contribution of the pure interaction effect between $\mathbf{p}_k$ and $\mathbf{p}_l$ to the total variance of the output variable $\mathbf{y}_j$ at time t_i, and so on. The sensitivity indices calculated in Eqs. (4)–(6) are in the range [0,1]. In particular, if the kth parameter of the model has a small effect on the output variable $\mathbf{y}_j$ at time t_i, $S_j^k(t_i)$ will be close to zero, whereas it will take a value closer to one if $\mathbf{p}_k$ has a strong effect on $\mathbf{y}_j(t_i)$. Guided by the values of the sensitivity indices thus calculated, the most sensitive model parameters can be selected, i.e., the parameters that account for most of the variability of model outputs.

Optimization Procedure

For $j = 1,\ldots,m$, m being the number of sampled compartments, define ϕ_j as that part of the objective function relative to the jth sampled compartment:

$$\phi_j(\mathbf{p}) = \sum_{i=1}^{n_j} \omega(\epsilon_{ij}, \mathbf{p}).$$

Let $\mathbf{p}^{(j)}$ be the p_j sensitive parameters on ϕ_j, identified from previous SA. Define the sets $E_j = \{\mathbf{p}^{(j)}\}$, and the intervals I_j of possible values for these parameters (from data available in the literature): determine the minimum ($m_1^{(j)}$ to $m_{p_j}^{(j)}$) and maximum ($M_1^{(j)}$ to $M_{p_j}^{(j)}$) possible values for each of these parameters, so as to

define the search intervals $I_j = [m_1^{(j)}, M_1^{(j)}] \times \cdots \times [m_{p_j}^{(j)}, M_{p_j}^{(j)}]$. Place the cardinals $c_j = |E_j|$ in increasing order and renumber them from the smallest to the largest so that $c_1 = min\ (c_j/j = 1,\ldots,m)$ and $c_m = max\ (c_j/j = 1,\ldots,m)$. Then, renumber the sets E_j and the intervals I_j. Thereafter, follow the successive steps below:

Step 1. Solve the optimization problem

$$\min_{I_1} \phi_1,$$

using a local optimization algorithm and starting from different initial points in the parameter space I_1. The non-varying parameters are set to any physiologically plausible value, that can be different for each initial guess. This procedure will provide a certain number of acceptable values for the model parameters included in E_1, which will be used to reduce the parameter space I_1 to a smaller search interval: the previous lower and upper bounds of each model parameter will be replaced by the minimum and maximum values newly found during the optimization process. Determine the minimum ($m_1^{(1)'}$ to $m_{p_1}^{(1)'}$) and maximum ($M_1^{(1)'}$ to $M_{p_1}^{(1)'}$) values obtained for each varying parameter, and let $m_1^{(1)} \leftarrow m_1^{(1)'}$, $m_{p_1}^{(1)} \leftarrow m_{p_1}^{(1)'}$, $M_1^{(1)} \leftarrow M_1^{(1)'}$ and $M_{p_1}^{(1)} \leftarrow M_{p_1}^{(1)'}$, so as to define the new search interval I_1.

Step k. ($2 \le k \le m$) Solve the optimization problem

$$\min_{I_1 \times \cdots \times I_k} \phi_1 + \cdots + \phi_k,$$

using a local optimization algorithm and starting from different initial points in the parameter space $I_1 \times \cdots \times I_k$. The non-varying parameters are set to any physiologically plausible value, that can be different for each initial guess. This will provide new minimum and maximum values for the model parameters included in $E_1 \cup \ldots \cup E_k$, allowing a reduction in the initial parameter space $I_1 \times \cdots \times I_k$ to a smaller search interval (as for I_1 in Step 1 of the algorithm).

Step m+1. Terminate by minimizing the complete objective function ϕ, using the previous multistart strategy and allowing all parameters to vary in the whole search space that has been significantly restricted during the precedent m steps of the algorithm for all sensitive parameters.

At intermediate step k of the algorithm, the idea is to find values for the sets of parameters $(\mathbf{p}^{(1)}, \ldots, \mathbf{p}^{(k)})$ that are not too far from the optimal values, so that they can be used during subsequent steps of the procedure as good starting points for the local optimizer. The last step ($m + 1$) of the algorithm consists in minimizing ϕ on the whole parameter space from starting points that are already close to their optimal values, so that any local optimizer associated with a multistart strategy will be successful in its optimization process. Various types of local search methods could be chosen here as an optimizer: a gradient descent method,[15] a Levenberg-Marquardt method, particularly suitable in least squares problems,[26,28] or a direct search method, which never calculates or approximates partial derivatives of the objective function with respect to model parameters.[34] The latter method should be preferred when information on the gradient of the objective function is either unavailable (no simple analytical formula) or unreliable (inaccurate calculations by finite differences). Because this concerns most complex compartmental models, we suggest use of a simplex, derivative-free algorithm, which has been presented in full detail elsewhere.[23,25]

As other local or global optimization techniques, the proposed methodology does not guarantee convergence to the global optimum. It will however be proven in the next section to be efficient to estimate the parameters of a complex linear compartmental model, for which the problem of numerical identification has already been shown to be extremely difficult to solve when using classic optimization techniques.[20]

APPLICATION

In this study, the algorithm has been applied to estimating the parameter values of a linear 13-compartment model describing the inter-organ postprandial metabolism of dietary N in humans (Fig. 1).[18] Although this selection reflects the specific interests of the authors, this model is one of the most complex linear compartmental models proposed in the field of human protein metabolism, so that the difficulties encountered when estimating its parameter values should be representative of those occurring when identifying a simpler one. Moreover, development of this model was first intended to use the experimental data available in some accessible pools in order to simulate the evolution of other compartments that could not be observed experimentally, despite their metabolic and nutritional importance. In such a context, it was truly important to find the exact global optimum of the inverse problem, and not a vicinity, i.e., some values—biologically acceptable or not—enabling the model satisfactorily to reproduce the measured responses. In this section, we demonstrate the effectiveness of our SA approach in solving the inverse problem using both simulated and real data.

Compartmental Model of Dietary N Metabolism in Humans

The linear compartmental model presented in Fig. 1 was developed from clinical data obtained in the

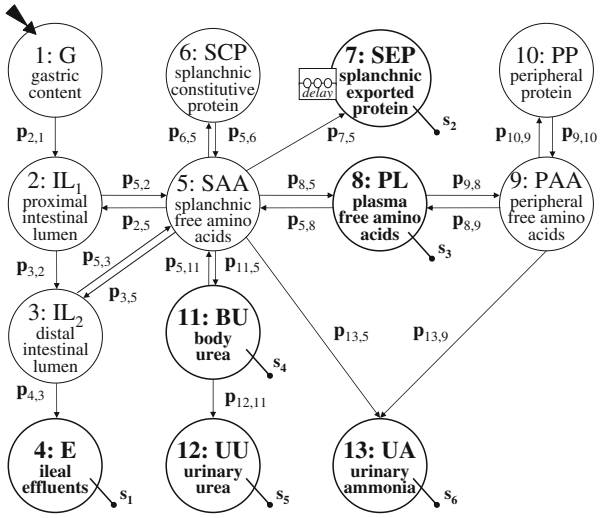

FIGURE 1. Linear 13-compartment model describing the postprandial distribution of dietary nitrogen (N) in humans, developed from clinical data obtained in the intestine, blood and urine of healthy adults after the ingestion of a [^{15}N]-labeled protein meal.[32] Circles indicate compartments representing kinetically distinct pools of dietary N; rectangle is a delay component; arrows between compartments represent transfer pathways, and numbers by the arrows indicate transfer rate constants ($\mathbf{p}_{i,j}$, fraction of dietary N in compartment *j* transferred to compartment *i* per min). Bullets indicate those compartments that were sampled (in bold): samples $s_1 - s_6$ represent cumulative ileal effluents (E), splanchnic exported proteins (SEP), plasma free amino acids (PL), body urea (BU), and cumulative urinary urea (UU) and ammonia (UA), respectively. Dietary N is considered to enter the first compartment (G: gastric content) as a bolus, and is then transferred through other compartments along the 21 transfer pathways at their corresponding transfer rates.

intestine, blood and urine of healthy adults after the ingestion of a [^{15}N]-labeled protein meal.[32] The model integrates 13 compartments representing the amount of dietary N present at a particular location in the body (e.g., dietary N present in the proximal intestinal lumen), or in a specific metabolic form (e.g., dietary N present in body urea), or a combination of both (e.g., dietary N present in a free AA form in the splanchnic area). The model also includes 21 metabolic pathways, each of them being characterized by a constant diffusion parameter $\mathbf{p}_{j_1,j_2}$ representing the constant fraction of dietary N transferred from compartment j_2 to compartment j_1 per unit of time. Of the 13 model compartments, only six were experimentally monitored: cumulative ileal effluents (E), splanchnic exported proteins (SEP), plasma free amino acids (PL), body urea (BU), cumulative urinary urea (UU), and cumulative urinary ammonia (UA). Compartments SEP, PL, BU, UU and UA were sampled every hour for 8 h, whereas compartment E was observed only once at 8 h. The other model compartments (G: gastric content; IL_1: proximal intestinal lumen; IL_2: distal intestinal lumen; SAA: splanchnic free amino acids; SCP: splanchnic constitutive proteins; PAA: peripheral free amino acids; and PP: peripheral proteins) were not accessible to experimental measurement for technical and/or ethical reasons.

The model is represented by Eqs. (1) and (2) previously described in section "System Model and Inverse Problem". $\mathbf{x}$ represents the 13-vector of model compartments, $\mathbf{x} = (G, IL_1, IL_2, E, SAA, SCP, SEP, PL, PAA, PP, BU, UU, UA)^t$, and A the 13×13 compartmental matrix whose entries are zeros or linear combinations of model parameters. No material entered the system from outside during the experiment ($B = 0$; $\forall t, \mathbf{u}(t) = 0$), except for the protein load from the meal represented by non-zero initial conditions ($\mathbf{x}(0) = (100,0,\ldots,0)^t$, all compartments being empty at time zero except the stomach, which received 100% of ingested N). Similarly, $\mathbf{y}$ represents the 6-vector of the six sampled compartments, $\mathbf{y} = (E, SEP, PL, BU, UU, UA)^t$, and C is a 6×13 output matrix whose entries are all zeros except those corresponding to the sampled compartments (equal to 1).

The global *a priori* identifiability of the model was tested using a new algorithm based on concepts of differential algebra and computer algebra tools,[1] to ensure that all model parameters had a unique solution in the ideal context of an error-free model structure and error-free experimental measurements.[2,7] The model was proven to be uniquely identifiable on the two conditions that: (i) the transfer rates from IL_1 to IL_2 and IL_2 to E were equal ($\mathbf{p}_{3,2} = \mathbf{p}_{4,3}$), and (ii) the transfer rates from SAA to IL_1 and SAA to IL_2 were also equal ($\mathbf{p}_{2,5} = \mathbf{p}_{3,5}$), these two conditions being consistent with our current knowledge of the physiological system. Consequently, there were 19 rather than 21 parameters to estimate. Model equations were integrated using an ODE solver designed for stiff problems. The objective function chosen to measure the quality of fit between model predictions and experimental data was the Logarithm of the Likelihood Function (LLF). Experimental errors were supposed to be normally distributed, with a mean of zero and a variance of σ_{ij}^2 for compartment j at time t_i. In this context of clinical experiments in humans, variances in measurement errors could not be determined experimentally, requiring postulation of an error model. The model chosen to represent the variance of measurement errors was that of Eq. (3). All the procedures described in section "Parameter Estimation Method" were implemented under Matlab 6.1 (The Mathworks Inc.).

Sensitivity Analysis

SA was carried out to identify model parameters with the greatest influence on each sampled compartment, according to the methodology suggested in section "Prior SA":

(1) We first determined the possible range of variation of each model parameter from current knowledge on the physiological system. The parameters were log-transformed to ensure an even distribution over the parameter space (Table 1);
(2) Then, 10,000 samples of parameter sets were generated from within the parameter space;
(3) For each of these samples, model outputs were computed by integrating the ODEs at each sampling time, i.e., SEP, PL, BU, UU and UA were evaluated every hour for 8 h and E was calculated at 8 h;
(4) Then, the total variance of each model output was computed at each sampling time, as well as the first-order sensitivity index of each model parameter, as described above.

Step 4 was performed 100 times in order to obtain 100 estimates of the total variance of each model output and 100 estimates of the first-order sensitivity index of each model parameter. They were averaged to obtain final estimates of the total variance and first-order sensitivity indices. Their accuracy was evaluated by computing standard deviation and 95% confidence interval, and was judged highly satisfactory. Because the sum of the first-order sensitivity indices computed at each sampling time for each sampled compartment always exceeded 90%, we did not calculate higher-order sensitivity indices. For each sampled compartment, the maximum-over-time first-order sensitivity indices were calculated (Table 2) and used to select the sensitive parameters: a model output was considered to be sensitive to a model parameter when the corresponding maximum-over-time sensitivity index exceeded 10%. Sampled compartments were then ordered according to their number of sensitive parameters. As indicated in Table 2, compartments E and SEP were sensitive mainly to four parameters, compartments BU and UU to five parameters and compartments PL and UA to six parameters. Compartments with the same number of sensitive parameters were then ordered according to their number of parameters having a non-zero contribution to the total variance.

Compartment E ($j = 1$) was thus selected as the compartment with the fewest sensitive parameters, followed by compartments SEP ($j = 2$), BU ($j = 3$), UU ($j = 4$), PL ($j = 5$) and UA ($j = 6$), in that order. As shown in Table 2, only four parameters ($\mathbf{p}_{2,1}$, $\mathbf{p}_{4,3}$, $\mathbf{p}_{5,2}$ and $\mathbf{p}_{5,3}$) made a contribution greater than 10% to the total output variance for compartment E. Thus, the

TABLE 1. Range of variation of the parameters of the model presented in Fig. 1.

Parameters	Lower bound	Upper bound
$\mathbf{p}_{2,1}$	−3	−1
$\mathbf{p}_{2,5}$	−4	−2
$\mathbf{p}_{4,3}$	−4	−1
$\mathbf{p}_{5,2}$	−3	−1
$\mathbf{p}_{5,3}$	−4	−1
$\mathbf{p}_{5,6}$	−5	−3
$\mathbf{p}_{5,8}$	−5	−2
$\mathbf{p}_{5,11}$	−4	−2
$\mathbf{p}_{6,5}$	−3	−1
$\mathbf{p}_{7,5}$	−3	−1
$\mathbf{p}_{8,5}$	−3	−1
$\mathbf{p}_{8,9}$	−3	−1
$\mathbf{p}_{9,8}$	−2	0
$\mathbf{p}_{9,10}$	−5	−3
$\mathbf{p}_{10,9}$	−3	−1
$\mathbf{p}_{11,5}$	−3	−1
$\mathbf{p}_{12,11}$	−4	−2
$\mathbf{p}_{13,5}$	−5	−2
$\mathbf{p}_{13,9}$	−5	−3

Only 19 on the 21 model parameters are reported in this table because of the equality constraints: $\mathbf{p}_{3,2} = \mathbf{p}_{4,3}$ and $\mathbf{p}_{2,5} = \mathbf{p}_{3,5}$. The values are over a logarithmic scale.

TABLE 2. Maximum-over-time percentage contributions of the parameters to the total variance in the sampled compartments of the model presented in Fig. 1.

	Output variables (%)					
Parameters	E	SEP	BU	UU	PL	UA
$\mathbf{p}_{2,1}$	15	37	42	33	22	38
$\mathbf{p}_{2,5}$	0	0	0	0	0	0
$\mathbf{p}_{4,3}$	79	11	11	11	4	11
$\mathbf{p}_{5,2}$	11	31	35	29	18	32
$\mathbf{p}_{5,3}$	11	3	4	2	1	3
$\mathbf{p}_{5,6}$	0	0	0	0	0	0
$\mathbf{p}_{5,8}$	0	0	0	0	0	0
$\mathbf{p}_{5,11}$	0	2	17	4	1	1
$\mathbf{p}_{6,5}$	0	9	9	6	3	9
$\mathbf{p}_{7,5}$	0	54	0	0	0	1
$\mathbf{p}_{8,5}$	0	6	7	4	20	6
$\mathbf{p}_{8,9}$	0	0	0	0	15	0
$\mathbf{p}_{9,8}$	0	0	0	0	46	0
$\mathbf{p}_{9,10}$	0	0	0	0	0	0
$\mathbf{p}_{10,9}$	0	0	0	0	18	16
$\mathbf{p}_{11,5}$	0	6	40	24	2	6
$\mathbf{p}_{12,11}$	0	0	9	31	0	0
$\mathbf{p}_{13,5}$	0	0	0	0	0	23
$\mathbf{p}_{13,9}$	0	0	0	0	0	13

Only 19 on the 21 model parameters are reported in this table because of the equality constraints: $\mathbf{p}_{3,2} = \mathbf{p}_{4,3}$ and $\mathbf{p}_{2,5} = \mathbf{p}_{3,5}$.

first step of the optimization algorithm consisted in fitting E by allowing only the parameter set $\mathbf{p}^{(1)} = (\mathbf{p}_{2,1}, \mathbf{p}_{4,3}, \mathbf{p}_{5,2}, \mathbf{p}_{5,3})$ to vary, the other parameters being set at physiologically plausible values (for example: $\mathbf{p}_{6,5}$, $\mathbf{p}_{7,5}$, $\mathbf{p}_{8,5}$ and $\mathbf{p}_{11,5}$ were all set at 10^{-2} min^{-1}, since they were known to lie within the interval $[10^{-3},10^{-1}]$). The second step of the algorithm consisted in fitting compartments E and SEP together by allowing $\mathbf{p}^{(1)}$ to vary (within a search space that had been significantly restricted from the first step of the algorithm), in addition to $\mathbf{p}^{(2)} = (\mathbf{p}_{7,5})$, which was the only additional parameter with a significant influence on SEP. This cascade of nested optimization processes continued following the steps described in Table 3, and finally ended when all sampled compartments were fitted simultaneously by variations to all model parameters. Besides, it can be seen from Table 2 that the parameters $\mathbf{p}_{2,5}$, $\mathbf{p}_{5,6}$, $\mathbf{p}_{5,8}$, and $\mathbf{p}_{9,10}$ have very little influence on the model responses. This suggests that these parameters may not be identified with good quality, by comparison with the parameters that can drive significant variations to the system outputs.

TABLE 3. Steps in the optimization procedure used to estimate parameters for the model presented in Fig. 1.

Steps	Notation	Characterization
1	$\phi = \phi_1$ $\mathbf{p} = \mathbf{p}^{(1)}$	$\phi_1 = \phi_E$ $\mathbf{p}^{(1)} = (\mathbf{p}_{2,1}, \mathbf{p}_{4,3}, \mathbf{p}_{5,2}, \mathbf{p}_{5,3})$
2	$\phi = \phi_1 + \phi_2$ $\mathbf{p} = (\mathbf{p}^{(1)},\mathbf{p}^{(2)})$	$\phi_2 = \phi_{SEP}$ $\mathbf{p}^{(2)} = (\mathbf{p}_{7,5})$
3	$\phi = \phi_1 + \cdots + \phi_3$ $\mathbf{p} = (\mathbf{p}^{(1)},\ldots,\mathbf{p}^{(3)})$	$\phi_3 = \phi_{BU}$ $x^{(3)} = (\mathbf{p}_{5,11}, \mathbf{p}_{11,5})$
4	$\phi = \phi_1 + \cdots + \phi_4$ $\mathbf{p} = (\mathbf{p}^{(1)},\ldots, \mathbf{p}^{(4)})$	$\phi_4 = \phi_{UU}$ $\mathbf{p}^{(4)} = (\mathbf{p}_{12,11})$
5	$\phi = \phi_1 + \cdots + \phi_5$ $\mathbf{p} = (\mathbf{p}^{(1)},\ldots, \mathbf{p}^{(5)})$	$\phi_5 = \phi_{PL}$ $\mathbf{p}^{(5)} = (\mathbf{p}_{8,5}, \mathbf{p}_{8,9}, \mathbf{p}_{9,8}, \mathbf{p}_{10,9})$
6	$\phi = \phi_1 + \cdots + \phi_6$ $\mathbf{p} = (\mathbf{p}^{(1)},\ldots, \mathbf{p}^{(6)})$	$\phi_6 = \phi_{UA}$ $\mathbf{p}^{(6)} = (\mathbf{p}_{13,5}, \mathbf{p}_{13,9})$
7	$\phi = \phi_1 + \cdots + \phi_6$	All model parameters

Application to Simulated Data

Our algorithm was first validated in the theoretical context of simulated data, either noise-free or with a known and controlled noise component, before being applied within a more realistic framework of clinical, noisy data. First of all, we simulated "noise-free" data (d_0) in the six sampled compartments at each sampling time (every hour for 8 h for compartments SEP, PL, BU, UU and UA, and at 8 h for compartment E), using a given set of parameter values (Table 4), previously found to be physiologically plausible.[18] We

TABLE 4. Comparison between true and estimated parameter values of the model presented in Fig. 1 when "noise-free" data (d_0), "low homogeneous noise" data (d_1), "high homogeneous noise" data (d_2), and "high heterogeneous noise" data (d_3) were injected into the objective function (Logarithm of the Likelihood Function, LLF).

	Parameter values (min^{-1})					Relative errors (%) Between true and estimated values			
		Estimated							
	True	From d_0	From d_1	From d_2	From d_3	d_0	d_1	d_2	d_3
$\mathbf{p}_{2,1}$	$1.00017 \cdot 10^{-2}$	$1.00018 \cdot 10^{-2}$	$1.00337 \cdot 10^{-2}$	$9.54457 \cdot 10^{-3}$	$1.29521 \cdot 10^{-2}$	$1.4 \cdot 10^{-3}$	0.3	4.6	29.5
$\mathbf{p}_{2,5}$	$1.79097 \cdot 10^{-3}$	$1.79033 \cdot 10^{-3}$	$1.20251 \cdot 10^{-3}$	$2.40709 \cdot 10^{-3}$	$2.18000 \cdot 10^{-3}$	$3.5 \cdot 10^{-2}$	32.9	34.4	21.7
$\mathbf{p}_{4,3}$	$5.45349 \cdot 10^{-3}$	$5.45346 \cdot 10^{-3}$	$5.46366 \cdot 10^{-3}$	$5.55255 \cdot 10^{-3}$	$4.78740 \cdot 10^{-3}$	$4.7 \cdot 10^{-4}$	0.2	1.8	12.2
$\mathbf{p}_{5,2}$	$1.23204 \cdot 10^{-2}$	$1.23203 \cdot 10^{-2}$	$1.21485 \cdot 10^{-2}$	$1.26723 \cdot 10^{-2}$	$1.03402 \cdot 10^{-2}$	$9.7 \cdot 10^{-4}$	1.4	2.9	16.1
$\mathbf{p}_{5,3}$	$1.28991 \cdot 10^{-2}$	$1.28988 \cdot 10^{-2}$	$1.28132 \cdot 10^{-2}$	$1.31822 \cdot 10^{-2}$	$1.23906 \cdot 10^{-2}$	$2.1 \cdot 10^{-3}$	0.7	2.2	3.9
$\mathbf{p}_{5,6}$	$1.38298 \cdot 10^{-4}$	$1.38302 \cdot 10^{-4}$	$1.32501 \cdot 10^{-4}$	$1.22296 \cdot 10^{-4}$	$1.48000 \cdot 10^{-4}$	$3.1 \cdot 10^{-3}$	4.2	11.6	7.0
$\mathbf{p}_{5,8}$	$4.83722 \cdot 10^{-4}$	$4.89490 \cdot 10^{-4}$	$2.55002 \cdot 10^{-4}$	$2.65055 \cdot 10^{-4}$	$4.14045 \cdot 10^{-4}$	1.2	47.3	45.2	14.4
$\mathbf{p}_{5,11}$	$1.30912 \cdot 10^{-3}$	$1.30912 \cdot 10^{-3}$	$1.30375 \cdot 10^{-3}$	$1.17113 \cdot 10^{-3}$	$1.11241 \cdot 10^{-3}$	$2.9 \cdot 10^{-4}$	0.4	10.5	15.0
$\mathbf{p}_{6,5}$	$5.71715 \cdot 10^{-2}$	$5.71737 \cdot 10^{-2}$	$5.81467 \cdot 10^{-2}$	$6.27020 \cdot 10^{-2}$	$4.70710 \cdot 10^{-2}$	$3.8 \cdot 10^{-3}$	1.7	9.7	17.7
$\mathbf{p}_{7,5}$	$1.25592 \cdot 10^{-2}$	$1.25592 \cdot 10^{-2}$	$1.27290 \cdot 10^{-2}$	$1.36554 \cdot 10^{-2}$	$1.07041 \cdot 10^{-2}$	$2.3 \cdot 10^{-4}$	1.4	8.7	14.8
$\mathbf{p}_{8,5}$	$3.51054 \cdot 10^{-2}$	$3.51055 \cdot 10^{-2}$	$3.52533 \cdot 10^{-2}$	$3.58606 \cdot 10^{-2}$	$3.57061 \cdot 10^{-2}$	$2.2 \cdot 10^{-4}$	0.4	2.2	1.7
$\mathbf{p}_{8,9}$	$9.54823 \cdot 10^{-3}$	$9.54861 \cdot 10^{-3}$	$9.07060 \cdot 10^{-3}$	$8.40858 \cdot 10^{-3}$	$1.04917 \cdot 10^{-2}$	$4.0 \cdot 10^{-3}$	5.0	11.9	9.9
$\mathbf{p}_{9,8}$	$1.72683 \cdot 10^{-1}$	$1.72678 \cdot 10^{-1}$	$1.69293 \cdot 10^{-1}$	$1.57559 \cdot 10^{-1}$	$2.14201 \cdot 10^{-1}$	$2.6 \cdot 10^{-3}$	2.0	8.8	24.0
$\mathbf{p}_{9,10}$	$2.01746 \cdot 10^{-4}$	$2.01751 \cdot 10^{-4}$	$1.20251 \cdot 10^{-4}$	$1.85586 \cdot 10^{-4}$	$2.51000 \cdot 10^{-4}$	$2.4 \cdot 10^{-3}$	40.4	8.0	24.4
$\mathbf{p}_{10,9}$	$1.05609 \cdot 10^{-2}$	$1.05605 \cdot 10^{-2}$	$9.92135 \cdot 10^{-3}$	$9.94375 \cdot 10^{-3}$	$9.69692 \cdot 10^{-3}$	$3.1 \cdot 10^{-3}$	6.1	5.8	8.2
$\mathbf{p}_{11,5}$	$4.32958 \cdot 10^{-2}$	$4.32957 \cdot 10^{-2}$	$4.38394 \cdot 10^{-2}$	$4.63397 \cdot 10^{-2}$	$3.64080 \cdot 10^{-2}$	$2.6 \cdot 10^{-4}$	1.3	7.0	15.9
$\mathbf{p}_{12,11}$	$1.46295 \cdot 10^{-3}$	$1.46295 \cdot 10^{-3}$	$1.46455 \cdot 10^{-3}$	$1.47691 \cdot 10^{-3}$	$1.44979 \cdot 10^{-3}$	$3.0 \cdot 10^{-5}$	0.1	1.0	0.9
$\mathbf{p}_{13,5}$	$3.71412 \cdot 10^{-4}$	$3.71414 \cdot 10^{-4}$	$3.78060 \cdot 10^{-4}$	$3.97463 \cdot 10^{-4}$	$2.83382 \cdot 10^{-4}$	$7.2 \cdot 10^{-4}$	1.8	7.0	23.7
$\mathbf{p}_{13,9}$	$3.93112 \cdot 10^{-5}$	$3.93111 \cdot 10^{-5}$	$3.86950 \cdot 10^{-5}$	$4.28946 \cdot 10^{-5}$	$4.57823 \cdot 10^{-5}$	$3.0 \cdot 10^{-4}$	1.6	9.1	16.5

Only 19 on the 21 model parameters are reported in this table because of the equality constraints: $\mathbf{p}_{3,2} = \mathbf{p}_{4,3}$ and $\mathbf{p}_{2,5} = \mathbf{p}_{3,5}$.
The CPU times reached 2–4 days depending on the data set considered.

also tested three different levels and types of variances for the measurement errors in accordance with the error model of Eq. (3):

(1) $\omega_j = 0.01$ and $\gamma_j = 2$ for all compartments (i.e., a constant relative error of 1%),
(2) $\omega_j = 0.05$ and $\gamma_j = 2$ for all compartments (i.e., a constant relative error of 5%),
(3) $\omega_j = 0.05$ and $\gamma_j = 2$ for compartments PL and BU (i.e., a constant relative error of 5%); $\omega_j = 0.1\ \mathbf{y}_j(t_1, \mathbf{p})$ and $\gamma_j = 0$ for compartments E, SEP, UU and UA (i.e., a constant absolute error of 10% of the value at 1 h of the considered compartment size).

In these three cases, the values chosen for ω_j and γ_j accounted for plausible error patterns of the real data sets. In the first two cases, a classical error pattern with a constant relative error ($\gamma_j = 2$) reaching 1 or 5% was chosen for all sampled compartments, because experimental errors were considered to be generally less than 5% of the considered compartment sizes. In the third case, a constant absolute error ($\gamma_j = 0$) reaching 10% of the value at 1h of the considered compartment size was chosen for compartments exhibiting bell-shaped curves. Indeed, this type of noise seemed to be more appropriate for cumulated data sets, for which the latest and highest data points were the most reliable. As a consequence, the third type of noise was more likely to represent correctly the set of data obtained from the clinical investigation. Finally, four sets of simulated data were available to us: one set of noise-free data (d_0), and three different sets of noisy data, referred to as "low homogeneous noise" data (d_1), "high homogeneous noise" data (d_2) and "high heterogeneous noise" data (d_3).

These four sets of simulated data were used to test (and study the impact of noise on) the effectiveness of our parameter estimation method. The results reported in Table 4 show that our algorithm was able to find again the set of true parameter values used to simulate "noise-free" data, with a very high level of precision: the average and maximum relative errors between the 19 true and estimated parameter values were 0.07 and 1.2%, respectively, with all model parameters (except $\mathbf{p}_{5,8}$) being estimated with a precision of less than 0.05%. The precision of parameter estimates was also evaluated by calculating the standard deviation of each parameter value from the inverse of the Fisher information matrix,[10] expressed in terms of a percent fractional standard deviation or coefficient of variation (CV). The smaller the CV value, the better was the estimated value of the parameter,[16] and parameter values with a CV inferior to 50% are usually judged to be adequately estimated.[30] As shown in Fig. 2, all parameters were estimated with very good precision for data d_0, since the highest CV for the main model parameters was inferior to 5%.

When noisy data d_1, d_2, and d_3 were injected into the LLF, the algorithm was also able to find again the initial set of parameter values, and precision remained satisfactory, as reported in Table 4. The average relative errors between the 19 true and estimated parameter values were about 8, 10, and 15% for data d_1, d_2, and d_3, respectively. Furthermore, when omitting parameters $\mathbf{p}_{2,5}$, $\mathbf{p}_{5,8}$, and $\mathbf{p}_{9,10}$, the average relative errors on the 16 remaining parameters decreased to 2, 7, and 14% for data d_1, d_2, and d_3, respectively. This precision was found to be very satisfactory, since we could not hope that the method would be able to estimate model parameters with better precision than the experimental error introduced in the data. Besides, the CV obtained at the end of fitting processes also indicated that model parameters were estimated with good precision (Fig. 2), since the highest CV for the main model parameters were inferior to 15% for data d_1 and about 30% for data d_2 and d_3. The fact that parameters $\mathbf{p}_{2,5}$, $\mathbf{p}_{5,8}$, and $\mathbf{p}_{9,10}$ had only a slight influence on the six sampled compartments, and that they exhibited strong correlations with parameters $\mathbf{p}_{5,2}$, $\mathbf{p}_{8,5}$, and $\mathbf{p}_{10,9}$, respectively, may explain the difficulties encountered when estimating their numerical values.

Application to Real Data

Once validated, the parameter estimation method was applied in the real context of clinical data obtained in healthy humans during the first 8 h after ingestion of a single [^{15}N]-labeled protein meal.[32] Model parameters, as well as parameters for the error model of Eq. (3), were estimated in order to generate the model predictions closest to experimental data. An interesting outcome of the error model was that it did not require any *a priori* knowledge on the pattern of measurement errors. It thus provided further information on the levels and types of experimental errors, the heteroscedasticity parameters allowing error variances to vary during parameter estimation between the two extreme cases of a constant absolute error and a constant relative error. This is of particular importance when dealing simultaneously with compartments with different scales and evolution profiles, as was the case in the application described here.

Our results were then validated to ensure that the algorithm provided the best possible fit between model predictions and experimental observations.[10] We thus successively evaluated the goodness-of-fit of model predictions (randomness and normality of residuals), the reliability of parameter estimates and the

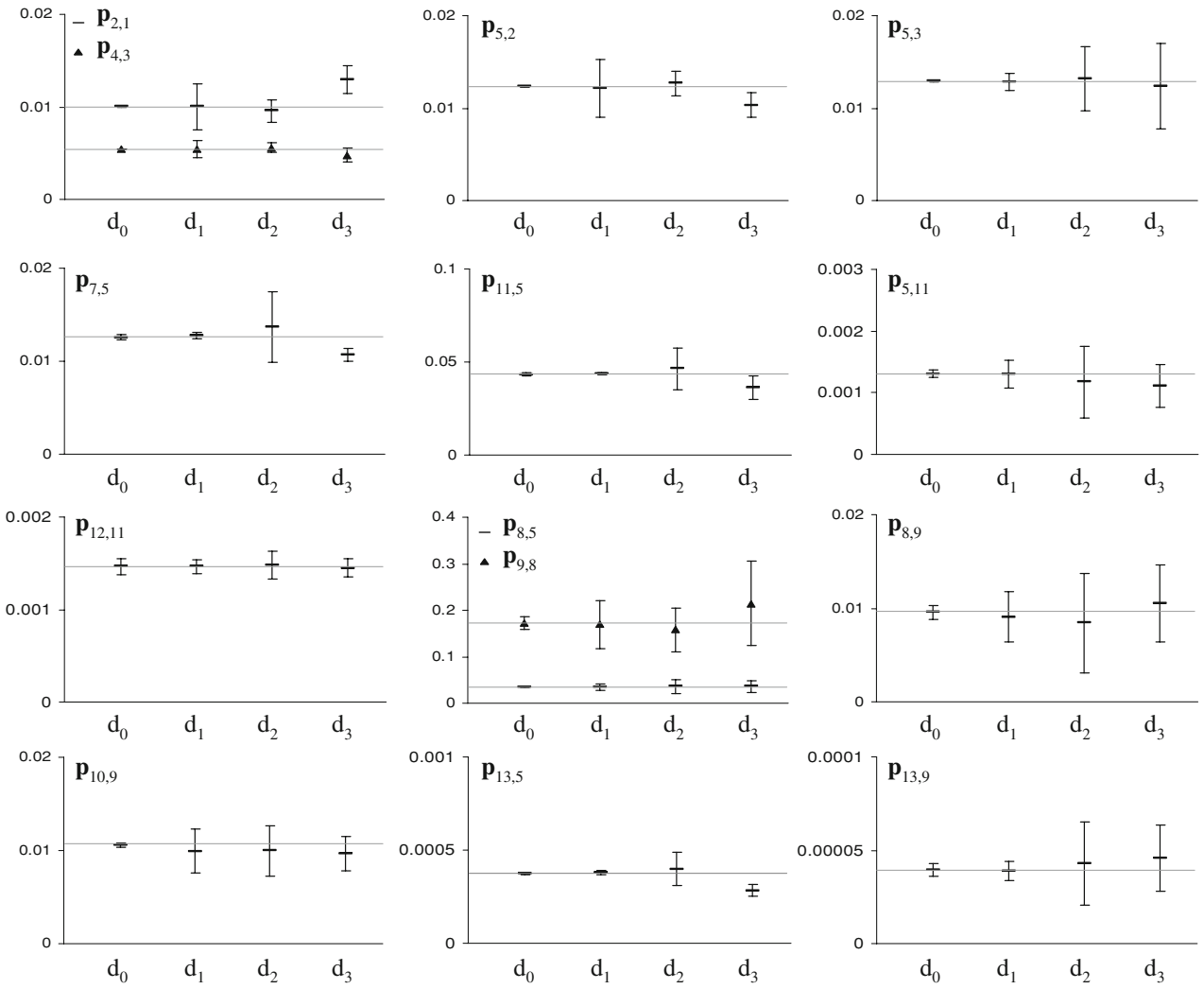

FIGURE 2. Estimated values (in min^{-1}) for sensitive parameters of the model presented in Fig. 1 when "noise-free" data (d_0), "low homogeneous noise" data (d_1), "high homogeneous noise" data (d_2), and "high heterogeneous noise" data (d_3) were injected into the objective function (Logarithm of the Likelihood Function, LLF). Horizontal lines represent the true parameter values used to simulate the data. Each estimated parameter is plotted by value ±2SD, with SD being calculated from the inverse of the Fisher information matrix.

consistency of model predictions with respect to corresponding findings in the literature.[9] Very good fits were achieved between model predictions and experimental data, as represented in Fig. 3. Moreover, most of the standardized residuals (i.e., residuals divided by the standard deviation of the experimental data) were randomly scattered around 0 and followed a normal distribution with a mean of 0 and a variance of 1 (i.e., 95% of standardized residuals lay within the range −1.96 to + 1.96), as shown in Fig. 4. In addition, the precision of parameter estimates was evaluated by calculating their CV: the mean CV was 4% and the highest CV reached 25%, indicating that model parameters were estimated with very good precision. Finally, the physiological relevance of model predictions (Fig. 5) was strongly supported by their consistency with respect to current data in the literature regarding gastrointestinal kinetics and the regional metabolism of dietary N, as summarized in Table 5.

CONCLUSION

Parameter estimation is already routinely applied to determining parameter values for compartmental models of biological systems.[7,24] However, parameter estimation still faces many mathematical challenges when several dynamic processes are concerned, as is generally the case for multi-compartmental models of complex physiological and nutritional systems.[37,45]

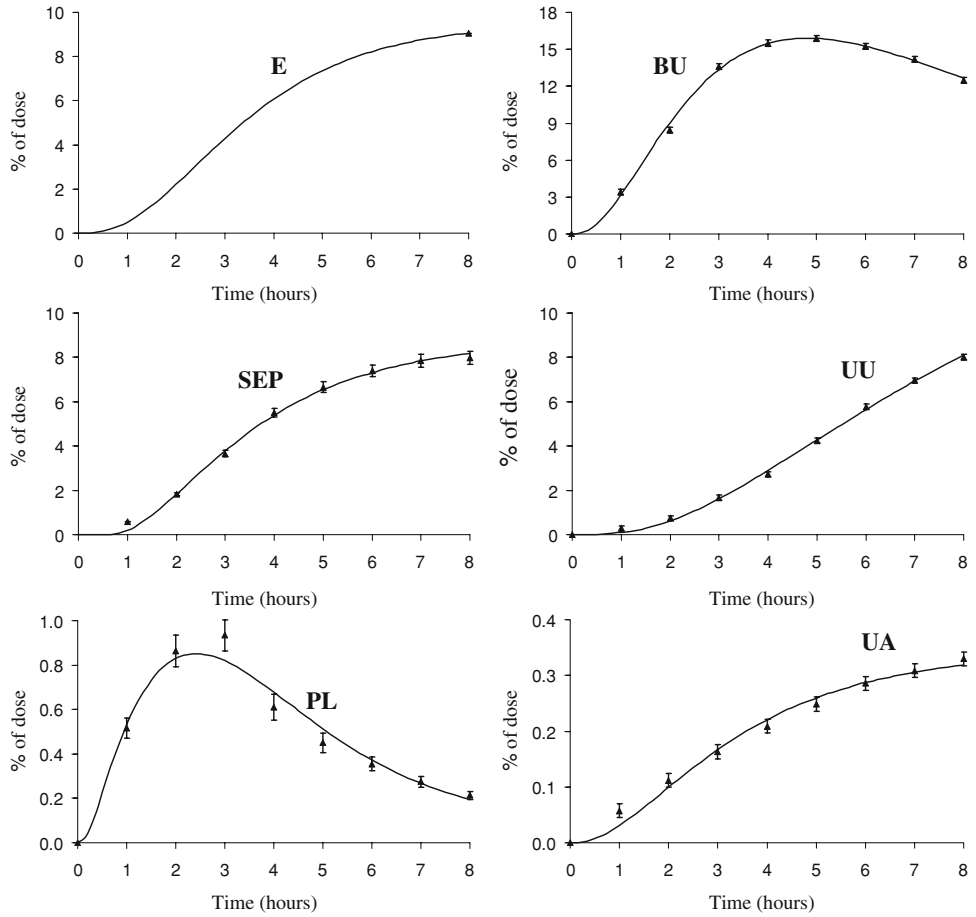

FIGURE 3. Observed vs. predicted values for the six sampled compartments of the model presented in Fig. 1. Lines: model predictions; Single points: experimental data. Each experimental datum is plotted by value ±SD, with the weighting scheme determined by the error model of Eq. (3). E: ileal effluents; SEP: splanchnic exported proteins; PL: plasma free amino acids; BU: body urea; UU: urinary urea; UA: urinary ammonia.

Determining global estimates, i.e., parameter values that allow the best possible fit between model predictions and experimental data for all sampled compartments, is one of these challenges, if not the most difficult.[40]

We describe here a strategy to estimate the parameters of linear compartmental models using a SA approach. A prior SA determines the relative influence of each parameter on the size of each sampled compartment, this information then being used to facilitate the parameter estimation process. The basic idea is to break down the large and hardly solvable problem (where all parameters vary simultaneously to optimize the whole objective function) into several simpler subproblems (where only a few parameters vary simultaneously to optimize a part of the objective function), which can easily be solved one after the other using a classic optimization technique. SA is already known to provide helpful information for model development, at the levels of both structural modeling and parameter estimation.[24,39] However, if SA has already been applied to reducing model complexity by removing or setting at their nominal values those parameters that contribute very little to variations in model outputs, our method is—as far as we know—the first that formalizes its use for identifying linear multi-compartmental models. Because it uses some *a priori* information available on the model behavior, the method presents analogies with the Bayesian approach that requires definition of a prior distribution for parameter values.[3] Compared to Fisherian parameter estimation techniques, Bayesian methods have already proved their efficiency in terms of both accuracy and precision.[35] However, assumption of this prior distribution requires a good knowledge of (or at least, sufficient information on) the kinetic properties of the system, which is not always the case, especially in humans. Our method can thus be seen as an alternative to classic parameter estimation techniques,

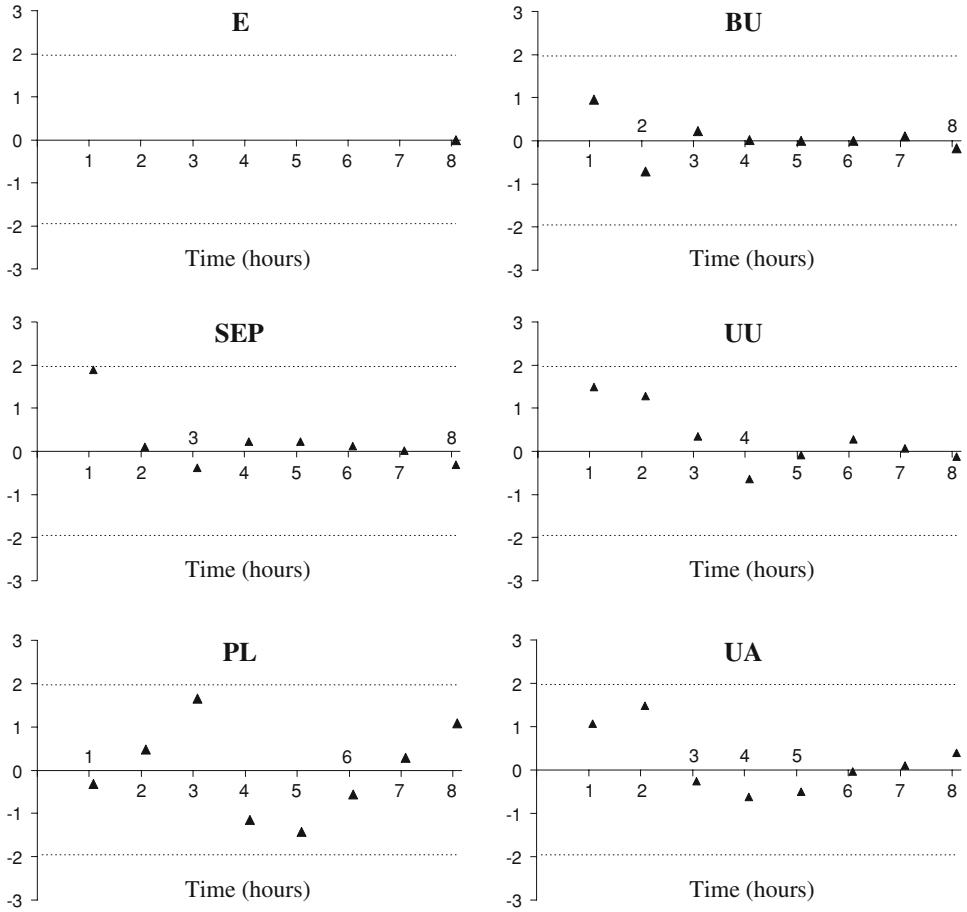

FIGURE 4. Weighted residuals at each sampling time for each sampled compartment of the model presented in Fig. 1. E: ileal effluents; SEP: splanchnic exported proteins; PL: plasma free amino acids; BU: body urea; UU: urinary urea; UA: urinary ammonia.

either when little information is available on parameter values, or when local optimization methods fail to find satisfactory estimates.

During this study, the usefulness of the proposed method was illustrated in the particular example of a linear 13-compartment model describing the postprandial metabolism of dietary N in humans. Identifying such a model constitutes a difficult optimization problem, since it is characterized by a large number of parameters (with little *a priori* information about their exact values) to be estimated from experimental data that are partial (only six sampled compartments), discrete (only eight data points per compartment) and noisy (unknown noise). We first validated the parameter estimation method in the ideal context of noise-free simulated data, and within a more realistic framework of noisy simulated data, the noise component being known and controlled. We then applied the method to experimental data obtained from clinical experiments and associated with unknown noise. Although no method would theoretically guarantee that the solution found at the end of the search procedure was indeed the global optimum, the *a posteriori* validation applied to our modeling results (goodness-of-fit, reliability of parameter estimates and physiological relevance of model predictions) ensured that the values obtained for model parameters were highly satisfactory. Nevertheless, there may be atypical models where our method will encounter difficulties. For instance, our algorithm may be unsuccessful in estimating the parameters of models with an insufficient number of sampled compartments, or models whose sampled compartments depend on very large and different sets of input parameters. However, other local and/or global optimization methods may not be much more efficient in such cases.

In conclusion, we hope that the methodology described here will be of value to the further development of linear multi-compartmental models representing the dynamic behavior of physiological and nutritional systems. Furthermore, its field of application may not be restricted to linear compartmental

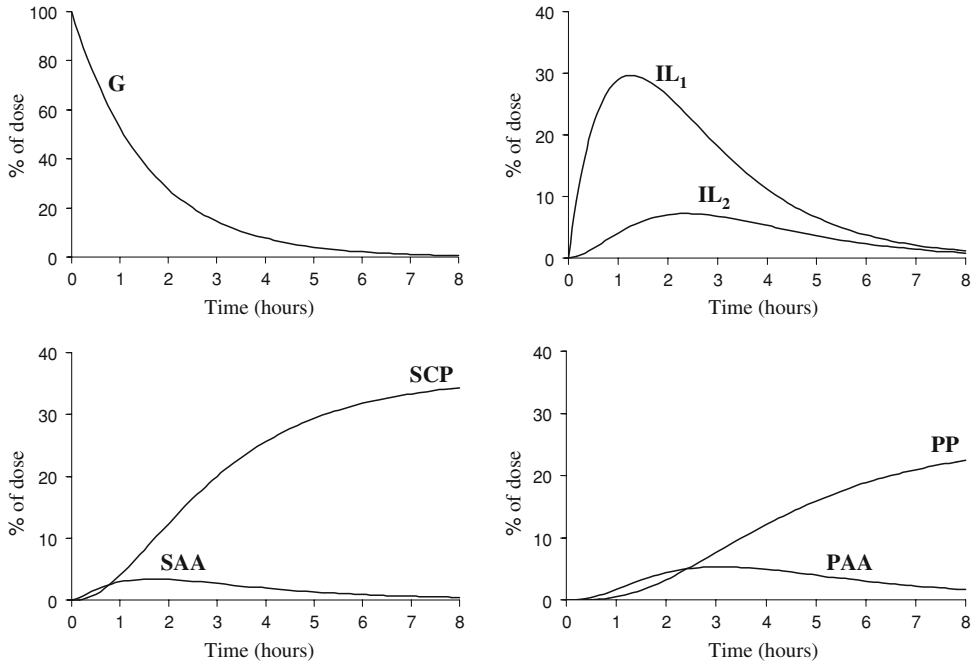

FIGURE 5. Predictions of the evolution kinetics for the non-sampled compartments in the model presented in Fig. 1. G: gastric content; IL_1: proximal intestinal lumen; IL_2: distal intestinal lumen; SAA: splanchnic free amino acids; SCP: splanchnic constitutive proteins; PAA: peripheral free amino acids; PP: peripheral proteins.

TABLE 5. External validation of the model presented in Fig. 1: comparison of model predictions with available findings in the literature.

Postprandial fate of dietary nitrogen		Model (% of ingested N)	Literature (% of ingested N)
Intestinal absorption	Proximal	66	69[a]
	Distal	19	20[a]
Splanchnic metabolism	Extraction	61	53–63[b]
	Anabolism	39	15–50[c]
Peripheral metabolism	Retention	22	18–50[d]
	Anabolism	19	18–38[e]

Model predictions vs. literature findings concerning the gastrointestinal kinetics and regional metabolism of dietary nitrogen (N) at 6 h after meal ingestion.
[a]Gausseres *et al.*[21]
[b]Biolo *et al.*,[5] Stoll *et al.*[43]
[c]Deutz *et al.*,[13] Stoll *et al.*[43]
[d]Capaldo *et al.*,[6] Elia *et al.*,[14] Morens *et al.*[33]
[e]Morens *et al.*[33]

models. It could be applied to any parametric model described by a set of ODE and partially observed at different sampling times during an experiment, e.g., nonlinear compartmental models or models represented by (nonlinear) power-law equations, including the Generalized Mass Action (GMA) and S-system models. Further investigations are required to confirm this point.

ACKNOWLEDGMENTS

The authors would like to thank M. P. Saccomani from the University of Padova (Padova, Italy) for having checked the *a priori* identifiability of the studied model, and for helpful and constructive discussions on the global *a priori* identifiability of compartmental models. The authors also thank E. Cancès from the École Nationale des Ponts et Chaussées (Paris, France) and P. Michel from the École Nationale Supérieure (Paris, France) for helpful and stimulating discussions.

REFERENCES

[1]Audoly, S., G. Bellu, L. D'Angiò, M. P. Saccomani, and C. Cobelli. Global identifiability of nonlinear models of biological systems. *IEEE Trans. Biomed. Eng*. 48:55–65, 2001.
[2]Audoly, S., L. D'Angio, M. P. Saccomani, and C. Cobelli. Global identifiability of linear compartmental models—a computer algebra algorithm. *IEEE Trans. Biomed. Eng*. 45:36–47, 1998.
[3]Bard, Y. Nonlinear Parameter Estimation. New York: Academic Press, 1974.
[4]Biolo, G., D. Chinkes, X. J. Zhang, and R. R. Wolfe. A new model to determine in vivo the relationship between amino acid transmembrane transport and protein kinetics in muscle. *JPEN J. Parenter. Enteral Nutr*. 16:305–315, 1992.

[5]Biolo, G., P. Tessari, S. Inchiostro, D. Bruttomesso, C. Fongher, L. Sabadin, M. G. Fratton, A. Valerio, and A. Tiengo. Leucine and phenylalanine kinetics during mixed meal ingestion: a multiple tracer approach. *Am. J. Physiol.* 262:E455–E463, 1992.

[6]Capaldo, B., A. Gastaldelli, S. Antoniello, M. Auletta, F. Pardo, D. Ciociaro, R. Guida, E. Ferrannini, and L. Sacca. Splanchnic and leg substrate exchange after ingestion of a natural mixed meal in humans. *Diabetes* 48:958–966, 1999.

[7]Carson, E. R., and C. Cobelli. Modeling Methodology for Physiology and Medicine. San Diego: Academic Press, 2001.

[8]Cayol, M., Y. Boirie, F. Rambourdin, J. Prugnaud, P. Gachon, B. Beaufrere, and C. Obled. Influence of protein intake on whole body and splanchnic leucine kinetics in humans. *Am. J. Physiol. Endocrinol. Metab.* 272:E584–E591, 1997.

[9]Cobelli, C., E. R. Carson, L. Finkelstein, and M. S. Leaning. Alidation of simple and complex models in physiology and medicine. *Am. J. Physiol.* 246:R259–266, 1984.

[10]Cobelli, C., and A. Caumo. Using what is accessible to measure that which is not: necessity of model of system. *Metabolism* 47:1009–1035, 1998.

[11]Cobelli, C., and D. M. Foster. Compartmental models: theory and practice using the SAAM II software system. *Adv. Exp. Med. Biol.* 445:79–101, 1998.

[12]Cobelli, C., M. P. Saccomani, P. Tessari, G. Biolo, L. Luzi, and D. E. Matthews. Compartmental model of leucine kinetics in humans. *Am. J. Physiol. Endocrinol. Metab.* 261:E539–E550, 1991.

[13]Deutz, N. E., M. J. Bruins, and P. B. Soeters. Infusion of soy and casein protein meals affects interorgan amino acid metabolism and urea kinetics differently in pigs. *J. Nutr.* 128:2435–2445, 1998.

[14]Elia, M., P. Folmer, A. Schlatmann, A. Goren, and S. Austin. Amino acid metabolism in muscle and in the whole body of man before and after ingestion of a single mixed meal. *Am. J. Clin. Nutr.* 49:1203–1210, 1989.

[15]Fletcher, R. Practical Methods of Optimization. Chichester: John Wiley and Sons, 1987.

[16]Foster, D. M. Developing and testing integrated multicompartmental models to describe a single-input multiple-output study using SAAM II software system. *Adv. Exp. Med. Biol.* 445:59–78, 1998.

[17]Fouillet, H., C. Bos, C. Gaudichon, and D. Tome. Approaches to quantifying protein metabolism in response to nutrient ingestion. *J. Nutr.* 132:3208S–3218S, 2002.

[18]Fouillet, H., C. Gaudichon, C. Bos, F. Mariotti, and D. Tome. Contribution of plasma proteins to splanchnic and total anabolic utilization of dietary nitrogen in humans. *Am. J. Physiol. Endocrinol. Metab.* 279:E88–E97, 2003.

[19]Fouillet, H., C. Gaudichon, F. Mariotti, C. Bos, J. F. Huneau, and D. Tome. Energy nutrients modulate the splanchnic sequestration of dietary nitrogen in humans: a compartmental analysis. *Am. J. Physiol. Endocrinol. Metab.* 281:E248–E260, 2001.

[20]Fouillet, H., C. Gaudichon, F. Mariotti, S. Mahe, P. Lescoat, J. F. Huneau, and D. Tome. Compartmental modeling of postprandial dietary nitrogen distribution in humans. *Am. J. Physiol. Endocrinol. Metab.* 279:E161–E175, 2000.

[21]Gausseres, N., S. Mahe, R. Benamouzig, C. Luengo, H. Drouet, J. Rautureau, and D. Tome. The gastro-ileal digestion of ^{15}N-labelled pea nitrogen in adult humans. *Br. J. Nutr.* 76:75–85, 1996.

[22]Guus, C., E. Boender, and H. E. Romeijn. Stochastics methods. In: Handbook of Global Optimization, edited by R. Horst and P. M. Pardalos. Dordrecht: Kluwer Academic Publishers, 1995, pp. 829–869.

[23]Humphrey, D. G., and J. R. Wilson. A revised simplex search procedure for stochastic simulation response surface optimization. *Informs J. Comput.* 12:272–283, 2000.

[24]Jacquez, J. A. Compartmental Analysis in Biology and Medicine. Ann Arbor, MI: BioMedware, 1996.

[25]Juillet, B., J. Salomon, D. Tome, and H. Fouillet. Development and calibration of a modeling tool for the analysis of clinical data in human nutrition. *ESAIM: Proc.* 14:124–155, 2005.

[26]Levenberg, K. A method for the solution of certain nonlinear problems in least squares. *Q. Appl. Math.* 2:164–168, 1944.

[27]Lowe, N. M., D. M. Shames, L. R. Woodhouse, J. S. Matel, R. Roehl, M. P. Saccomani, G. Toffolo, C. Cobelli, and J. C. King. A compartmental model of zinc metabolism in healthy women using oral and intravenous stable isotope tracers. *Am. J. Clin. Nutr.* 65:1810–1809, 1997.

[28]Marquardt, D. W. An algorithm for least squares estimation of nonlinear parameters. *SIAM J.* 11:431–441, 1963.

[29]Mendes, P., and D. B. Kell. Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. *Bioinformatics* 14:869–883, 1998.

[30]Miller, L. V., N. F. Krebs, and K. M. Hambidge. Human zinc metabolism: advances in the modeling of stable isotope data. *Adv. Exp. Med. Biol.* 445:253–269, 1998.

[31]Moles, C. G., P. Mendes, and J. R. Banga. Parameter estimation in biochemical pathways: a comparison of global optimization methods. *Genome Res.* 13:2467–2474, 2003.

[32]Morens, C., C. Bos, M. E. Pueyo, R. Benamouzig, N. Gausseres, C. Luengo, D. Tome, and C. Gaudichon. Increasing habitual protein intake accentuates differences in postprandial dietary nitrogen utilization between protein sources in humans. *J. Nutr.* 133:2733–2740, 2003.

[33]Morens, C., C. Gaudichon, C. C. Metges, G. Fromentin, A. Baglieri, P. C. Even, J. F. Huneau, and D. Tome. A high-protein meal exceeds anabolic and catabolic capacities in rats adapted to a normal protein diet. *J. Nutr.* 130:2312–2321, 2000.

[34]Nelder, J. A., and R. Mead. A simplex method for function minimization. *Comput. J.* 7:308–313, 1965.

[35]Pillonetto, G., G. Sparacino, and C. Cobelli. Numerical non-identifiability regions of the minimal model of glucose kinetics: superiority of Bayesian estimation. *Math. Biosci.* 184:53–67, 2003.

[36]Rabitz, H., Ö. F. Alis, J. Shorter, and K. Shim. Efficient input–output model representations. *Comput. Phys. Comm.* 117:11–20, 1999.

[37]Rodriguez-Fernandez, M., P. Mendes, and J. R. Banga. A hybrid approach for efficient and robust parameter estimation in biochemical pathways. *BioSystems* 83:248–265, 2006.

[38]Saccomani, M. P., R. C. Bonadonna, D. M. Bier, R. A. DeFronzo, and C. Cobelli. A model to measure insulin effects on glucose transport and phosphorylation in muscle: a three-tracer study. *Am. J. Physiol. Endocrinol. Metab.* 270:E170–E185, 1996.

[39]Saltelli, A., S. Tarantola, and F. Compologo. Sensitivity analysis as an ingredient of modelling. *Stat. Sci.* 15:377–395, 2000.
[40]Schittkowski, K. Numerical Data Fitting in Dynamical Systems—A Practical Introduction with Applications and Software. Kluwer Academic Publishers, 2002.
[41]Shyr, L. J., W. C. Griffith, and B. B. Boecker. An optimization strategy for a biokinetic model of inhaled radionuclides. *Fundam. Appl. Toxicol.* 16:423–434, 1991.
[42]Sobol', I. M. Sensitivity estimates for nonlinear mathematical models. *Math. Model. Comput. Exp.* 1:407–414, 1993.
[43]Stoll, B., D. G. Burrin, J. Henry, H. Yu, F. Jahoor, and P. J. Reeds. Dietary amino acids are the preferential source of hepatic protein synthesis in piglets. *J. Nutr.* 128:1517–1524, 1998.
[44]Tessari, P., S. Inchiostro, M. Zanetti, and R. Barazzoni. A model of skeletal muscle leucine kinetics measured across the human forearm. *Am. J. Physiol. Endocrinol. Metab.* 269:E127–E136, 1995.
[45]Tsai, K. Y., and F. S. Wang. Evolutionary optimization with data collocation for reverse engineering of biological networks. *Bioinformatics* 21:1180–1188, 2005.
[46]Twisk, J., D. L. Gillian-Daniel, A. Tebon, L. Wang, P. H. R. Barrett, and A. D. Attie. The role of the LDL receptor in apolipoprotein B secretion. *J. Clin. Invest.* 105:521–532, 2000.