Journal of Pharmacokinetics and Pharmacodynamics, Vol. 32, Nos. 5–6, December 2005 (© 2005)
DOI: 10.1007/s10928-005-0054-y

Lumping in Pharmacokinetics

Céline Brochot,[1,*] János Tóth,[2] and Frédéric Y. Bois[1]

Received May 17, 2005—Final August 2, 2005

Pharmacokinetic (PK) models simplify biological complexity by dividing the body into interconnected compartments. The time course of the chemical's amount (or concentration) in each compartment is then expressed as a system of ordinary differential equations. The complexity of the resulting system of equations can rapidly increase if a precise description of the organism is needed. However, difficulties arise when the PK model contains more variables and parameters than comfortable for mathematical and computational treatment. To overcome such difficulties, mathematical lumping methods are new and powerful tools. Such methods aim at reducing a differential system by aggregating several variables into one. Typically, the lumped model is still a differential equation system, whose variables are interpretable in terms of variables of the original system. In practice, the reduced model is usually required to satisfy some constraints. For example, it may be necessary to keep state variables of interest for prediction unlumped. To accommodate such constraints, constrained lumping methods have are also available. After presenting the theory, we study, here, through practical examples, the potential of such methods in toxicol/pharmacokinetics. As a tutorial, we first simplify a 2-compartment pharmacokinetic model by symbolic lumping. We then explore the reduction of a 6-compartment physiologically based pharmacokinetic model for 1,3-butadiene with numerical constrained lumping. The lumping methods presented here can be easily automated, and are applicable to first-order ordinary differential equation systems.

KEY WORDS: pharmacokinetics; physiologically-based pharmacokinetic model; unconstrained lumping; constrained lumping; 1,3-butadiene.

INTRODUCTION

Two kinds of pharmacokinetic (PK) models, or toxicokinetic (TK) models (for toxic compounds), are typically used to describe the absorption, distribution, metabolism, and elimination of chemicals as a function of time:

[1]INERIS, Institut National de l'Environnement Industriel et des Risques, Experimental Toxicology Unit, Parc Alata BP2, 60550 Verneuil en Halatte, France.
[2]Department of Mathematical Analysis, Budapest University of Technology and Economics, Hungary.
*To whom correspondence should be addressed. E-mail: Celine.Brochot@ineris.fr

1567-567X/05/1200-0719/0

data-based PK models (also referred as empirical PK models) and physiologically-based pharmacokinetic (PBPK) models (1). They simplify biological complexity by dividing the body into interconnected compartments. The time evolution of the chemical′s amount (or concentration) in each compartment is described by a system of ordinary differential equations. The dimension of such systems depends on model complexity. Typically, the finer the description of the biodistribution process, the higher the system dimension. Due to their pretense at describing anatomical and physiological structure, PBPK models have typically higher complexity and dimensionality than empirical PK models. However, difficulties arise when the PK model contains more variables and parameters than comfortable for mathematical and computational treatment. Indeed, parameter estimation, dosage design, and optimization are not easily handled with high-dimensional models.

Lumping techniques aim at reducing model dimensionality and complexity by aggregating several variables into one. The aggregation may concern different chemical species (sometimes unknown as in a combustion process) in the case of concomitant exposure to several agents for example (2,3). At first sight, one needs n PK models to describe the behavior of n substances. To simplify modeling and calculations, agents exhibiting the same physico–chemical and biodistribution properties can be lumped in a single "virtual species". Only one PK model is needed to model the kinetics of this new species. Another lumping approach, the one of interest in this paper, deals with the model's compartmental structure. In that case, several compartments are grouped into one. Usually, the lumped model is still an ordinary differential equations system with new variables, interpretable in terms of variables of the original system. So far, in PK, to that aim only semi-empirical methods have been proposed (4,5). Such procedures are applicable to PBPK models, but only compartments with nearly identical specifications and occupying equivalent positions in the system structure can be lumped together. For example, lumped tissues should have similar or close values for time constants, or should equilibrate very rapidly with each other. Hence, semi-empirical lumping processes always depend on parameter values, and should be re-evaluated for each (class of) substance(s).

Alternatively, several mathematical (exact or approximate) lumping methods have been proposed (6,7). These methods are only based on structure of differential equations systems, and are applicable to linear or nonlinear systems (6,8,9). Depending on its complexity, the system of interest can be treated either numerically or symbolically. Such methods have been originally developed in the fields of atmospheric or petroleum chemistry and combustion (10–12). They can provide effective solutions (demonstrated through past applications) to the need for a computational

approximation of a chemical mechanism (11,13,14). In practice, the reduced model is usually required to satisfy some restrictions. For example, it can be necessary to keep unlumped some state variables which have been experimentally measured (such as blood concentration), or which are of interest for prediction (for example, the quantity metabolized). To accommodate such restrictions, constrained lumping methods have also been developed (11,15,16). Overall, the choice of a lumping method depends essentially on the objectives of the study and on the model structure.

In this paper, we focus on linear PK models, i.e. corresponding to linear differential equation systems. We present mathematical lumping methods applicable to such models, together with examples. First, we briefly introduce the fundamentals of unconstrained and constrained lumping for linear differential equations systems (as a special case of general differential equations systems). Then, symbolic lumping is applied to a general 2-compartment model. Finally, a PBPK model for 1,3-butadiene (BD) biodistribution is treated by numerical constrained lumping.

METHODS

In this section, we present the constrained and unconstrained lumping methods for linear systems of first-order differential equations. For simplicity, some standard mathematical assumptions are omitted. More details can be found in Li and Rabitz (6,16).

PK Models and the Lumping Process Theory

A general definition of a PK model (of dimension n) can be given by

$$\frac{d\mathbf{y}(t)}{dt} = \mathbf{K}\mathbf{y}(t) + \mathbf{u}(t) \tag{1}$$

where $\mathbf{y}$ is the vector containing the n state variables (usually the amount or concentration of the chemical in a compartment), $\mathbf{u}$ the inflow vector (such as through inhalation) and $\mathbf{K}$ the matrix of coefficients (of dimension $n \times n$). For empirical compartmental PK models, elements of $\mathbf{K}$ are usually the transfer rate constants (or linear combinations of them) between the compartments. For PK models, some restrictions on the elements of $\mathbf{K}$ occur (17):

$$\begin{aligned} K_{ij} &\geq 0 \quad i \neq j; \quad j = 1 \ldots n \\ K_{ii} &\leq -\sum_{\substack{j=1 \\ j \neq i}}^{n} K_{ji} \quad i = 1 \ldots n \end{aligned} \tag{2}$$

To lump the system (Eq. 1) into a $\hat{n}$-dimensional system ($\hat{n} \leq n$), we introduce some notation. Let $\hat{\mathbf{y}}$ be the $\hat{n}$-dimensional vector containing the state variables of the new lumped system, and $\hat{\mathbf{u}}$ the new inflow vector. These two vectors are obtained by

$$\hat{\mathbf{y}}(t) = \mathbf{M}\mathbf{y}(t) \quad \text{and} \quad \hat{\mathbf{u}}(t) = \mathbf{M}\mathbf{u}(t) \tag{3}$$

where $\mathbf{M}$ is a constant matrix (of dimension $\hat{n} \times n$), called "lumping matrix". The system (Eq. 1) is considered to be lumpable if there exists a matrix $\hat{\mathbf{K}}$ (dimension $\hat{n} \times \hat{n}$) such that

$$\frac{d\hat{\mathbf{y}}(t)}{dt} = \hat{\mathbf{K}}\hat{\mathbf{y}}(t) + \hat{\mathbf{u}}(t) \tag{4}$$

Li and Rabitz (6) proved that such a matrix $\hat{\mathbf{K}}$ can be obtained as:

$$\hat{\mathbf{K}} = \mathbf{M}\mathbf{K}\overline{\mathbf{M}} \tag{5}$$

where $\overline{\mathbf{M}}$ is a generalized inverse of $\mathbf{M}$ (i.e., their matrix multiplication gives the identity matrix of dimension $\hat{n} \times \hat{n}$).

The task is now to construct the lumping matrices $\mathbf{M}$. In the following paragraphs, the construction of such lumping matrices is presented in the unconstrained and constrained cases.

Construction of the Lumping Matrix M

In the case of linear systems, the construction of the lumping matrix $\mathbf{M}$ is relatively straightforward (6). When the aggregation is exact, the subspace spanned by the rows of $\mathbf{M}$ should be $\mathbf{J}^{\mathrm{T}}(\mathbf{y})$-invariant, with $\mathbf{J}^{\mathrm{T}}(\mathbf{y})$ the transpose of the Jacobian matrix of $\mathbf{K}\mathbf{y}$. This latter can be expanded according to

$$\mathbf{J}^{\mathbf{T}}(\mathbf{y}) = \sum_{k=1}^{m} a_k(\mathbf{y})\mathbf{A}_k \tag{6}$$

where m is less than n^2, and the constant matrices $\mathbf{A}_k$ are viewed as a set of basis matrices of $\mathbf{J}^{\mathrm{T}}(\mathbf{y})$. For a unimolecular reaction system and in particular for linear PK systems, the Jacobian matrix is $\mathbf{K}$,

$$\mathbf{J}^{\mathbf{T}}(\mathbf{y}) = \mathbf{K}^{\mathbf{T}} \tag{7}$$

Therefore, we have to find spaces of dimension $\hat{n}$ that are invariant to $\mathbf{K}^{\mathrm{T}}$ in order to determine $\mathbf{M}$.

Unconstrained Lumping

Therefore, the rows of a $\hat{n}$-dimensional matrix **M** are composed by any of the $\hat{n}$ eigenvectors of the transpose of **K**. Several lumping matrices can be defined for the same system dimension. Indeed, $n!/(\hat{n}!(n-\hat{n})!)$ $\hat{n}$-dimensional lumping matrices can be constructed for a n-dimensional system with distinct eigenvalues. It is also possible to multiply (from the left) such lumping matrices by a nonsingular square matrix without changing their lumping property.

Theoretically, any first-order linear differential equations systems can be treated symbolically with this unconstraint lumping procedure. However, in practice, the calculation of eigenvalues and eigenvectors is not symbolically feasible for large dimensional matrices/models. In such cases, only numerical lumping can be performed.

Constrained Lumping

To accommodate restrictions required when dealing with practical examples, the lumping can be constrained by *a priori* specification of a part of the lumping matrix **M**. This latter is composed by two sub-matrices,

$$\mathbf{M} = \begin{pmatrix} \mathbf{M_G} \\ \mathbf{M_D} \end{pmatrix}$$

$\mathbf{M_G}$ given by the constraints and $\mathbf{M_D}$ to be determined. Under these constraints, exact lumping scheme may not exist. Therefore, the determination of constrained approximate lumping schemes is necessary. Li and Rabitz (15,16) have proposed the direct constrained approximate lumping (DCAL) method to determine the matrix **M**. This method consists in determining the base matrices $\mathbf{A_k}$ numerically (Eq. 6) by using values of **y** in a region of the n-dimensional composition space of interest. The methodology is exposed in Appendix A.

Since the base matrices $\mathbf{A_k}$ are determined numerically, this lumping method can only be applied when the model is fully described, i.e. model parameter values are known.

APPLICATIONS OF MATHEMATICAL LUMPING IN THE PK FIELD

In this section, we apply lumping methods to some PK models. First, a general 2-compartment model is simplified by symbolic unconstrained lumping. A 6-compartment PBPK model for 1,3-butadiene (BD) is then presented and simplified with numerical constrained lumping. The *Mathematica* software was used for all calculations.

Symbolic Lumping of a General Two-Compartment Model

Let us consider a general two-compartment model (Fig. 1),

$$\begin{aligned}\frac{dy_1(t)}{dt} &= -(k_{12}+k_{10})\,y_1(t) + k_{21}y_2(t) + k_{01}(t)\\ \frac{dy_2(t)}{dt} &= k_{12}y_1(t) - (k_{21}+k_{20})\,y_2(t) + k_{02}(t)\end{aligned} \tag{8}$$

All the rate coefficients are nonnegative real constants. The matrix **K** and the vector **u** are then given by

$$\mathbf{K} = \begin{pmatrix} -(k_{12}+k_{10}) & k_{21} \\ k_{12} & -(k_{21}+k_{20}) \end{pmatrix} \quad \mathbf{u}(t) = \begin{pmatrix} k_{01}(t) \\ k_{02}(t) \end{pmatrix} \tag{9}$$

All the rate coefficients are nonnegative real constants with the natural constraint that a is positive (this does not in fact restrict generality). This system is of the form of the Eq. 1, and the matrix **K** and the vector **u** were given in "Applications of Mathematical lumping in the PK field" section. λ_1 and λ_2, the two eigenvalues of $\mathbf{K}^{\mathrm{T}}$, are

$$\begin{aligned}\lambda_1 &= \frac{1}{2}\Big(-k_{12}-k_{21}-k_{10}-k_{20} \\ &\quad + \sqrt{(k_{12}+k_{21}+k_{10}+k_{20})^2 - 4\,(k_{21}k_{10}+k_{12}k_{20}+k_{10}k_{20})}\Big) \\ \lambda_2 &= \frac{1}{2}\Big(-k_{12}-k_{21}-k_{10}-k_{20} \\ &\quad - \sqrt{(k_{12}+k_{21}+k_{10}+k_{20})^2 - 4\,(k_{21}k_{10}+k_{12}k_{20}+k_{10}k_{20})}\Big)\end{aligned} \tag{10}$$

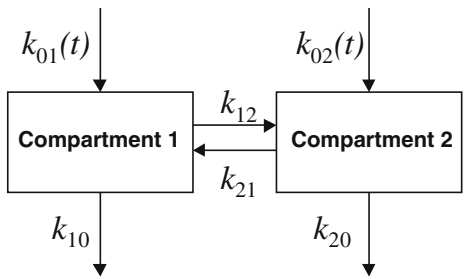

Fig. 1. Representation of the general 2-compartment model used. Arrows describe the exchange of material.

Following the unconstrained lumping method described in "Applications of Mathematical lumping in the PK field" section, the lumping matrix **M** is composed by the two eigenvectors of $\mathbf{K}^{\mathrm{T}}$:

$$s_1 = \begin{pmatrix} -k_{12} + k_{21} - k_{10} + k_{20} + \sqrt{(-k_{12} + k_{21} - k_{10} + k_{20})^2 + 4k_{12}k_{21}} \\ 2k_{21} \end{pmatrix}$$
$$s_2 = \begin{pmatrix} -k_{12} + k_{21} - k_{10} + k_{20} - \sqrt{(-k_{12} + k_{21} - k_{10} + k_{20})^2 + 4k_{12}k_{21}} \\ 2k_{21} \end{pmatrix} \tag{11}$$

Two one-dimensional lumping systems are therefore possible:

$$\begin{aligned} \frac{d\hat{\mathbf{y}}(t)}{dt} &= -\lambda_1 \hat{\mathbf{y}}(t) + s_1 \mathbf{u}(t) \\ \frac{d\hat{\mathbf{y}}(t)}{dt} &= -\lambda_2 \hat{\mathbf{y}}(t) + s_2 \mathbf{u}(t) \end{aligned} \tag{12}$$

The new variable is therefore a combination of the two original variables, as defined by Eq. 3,

$$\hat{y}(t) = s_i^T y(t) \tag{13}$$

where i is equal to 1 or 2, according to the selected lumping scheme. Let's take an example. Consider a 2-compartment model in which exchanges between the two compartments are equal, i.e. $k_{12} = k_{21}$, and rates going out of the compartments are also equal, i.e. $k_{10} = k_{20}$. Dividing each element by 2 k_{12}, the eigenvector s_1 of $\mathbf{K}^{\mathrm{T}}$ is (1; 1). We then obtain the new lumped system

$$\frac{d\hat{\mathbf{y}}(t)}{dt} = -\lambda_1 \hat{\mathrm{y}}(t) + k_{01}(t) + k_{02}(t) \tag{14}$$

with

$$\hat{\mathbf{y}}(t) = y_1(t) + y_2(t) \tag{15}$$

We have therefore reduced the 2-compartment model into a 1-compartment one, for which the new state variable is the sum of the two original ones.

Numerical Constrained Lumping of a Whole-Body PBPK Model for 1,3-Butadiene

In this section, we propose to simplify a 7-compartment whole-body PBPK model for BD. The BD is a chemical compound largely used in the production of plastics and synthetic rubber. Studies in rats and mice have demonstrated that its metabolites can cause cancer. Human BD exposure data are available and have been analyzed using population pharmacokinetic tools (18,19). Yet the computations required are heavy and would benefit from a simpler model with similar performance. For actual applications, we should make sure that the new lumped model is still able to answer the question(s) of interest. For this, we can apply constrained lumping methods. One constraint arise in our example: to evaluate BD toxicity, the total quantity of BD metabolized after a given exposure should still be quantified by the new model, and therefore should not be lumped. First, the full PBPK model is presented, lumped models are derived and effects of lumping on the results of metabolism predictions are studied.

PBPK Model for 1,3-Butadiene

The BD kinetics in the human body can be well described by a whole-body PBPK model, including seven compartments (blood, fat, liver, lungs, slowly perfused tissues—referred here to as "muscle"—, other rapidly perfused tissues—"viscera", and metabolism) and three sites of metabolism (liver, lungs and viscera) (20). The BD can enter the body by inhalation. We linearized Kohn and Melnick's model (20) by using first order metabolism terms, rather than Michaelis–Menten terms. The model is still relevant for the usually found low concentrations exposures (21). The model we used is therefore given by the following system of differential equations (of dimension 7):

$$
\begin{aligned}
\frac{dQ_B(t)}{dt} &= -F_C \times \left(\frac{Q_B(t)}{V_B} - \frac{Q_{Lu}(t)}{V_{Lu} \times PC_{Lu}}\right) - \sum_{i \in I} F_i \times \left(\frac{Q_B(t)}{V_B} - \frac{Q_i(t)}{V_i \times PC_i}\right) \\
\frac{dQ_F(t)}{dt} &= F_F \times \left(\frac{Q_B(t)}{V_B} - \frac{Q_F(t)}{V_F \times PC_F}\right) \\
\frac{dQ_M(t)}{dt} &= F_M \times \left(\frac{Q_B(t)}{V_B} - \frac{Q_M(t)}{V_M \times PC_M}\right) \\
\frac{dQ_{Li}(t)}{dt} &= F_{Li} \times \left(\frac{Q_B(t)}{V_B} - \frac{Q_{Li}(t)}{V_{Li} \times PC_{Li}}\right) - K_{Li} \times Q_{Li}(t) \\
\frac{dQ_{Lu}(t)}{dt} &= F_C \times \left(\frac{Q_B(t)}{V_B} - \frac{Q_{Lu}(t)}{V_{Lu} \times PC_{Lu}}\right) \\
&\quad + K_{Vent} \times \left(C_{Inh}(t) - \frac{Q_{Lu}(t)}{V_{Lu} \times PC_{Air}}\right) - K_{Lu} \times Q_{Lu}(t)
\end{aligned}
\tag{16}
$$

$$\frac{dQ_V(t)}{dt} = F_V \times \left(\frac{Q_B(t)}{V_B} - \frac{Q_V(t)}{V_V \times PC_V}\right) - K_V \times Q_V(t)$$
$$\frac{dQ_{Met}(t)}{dt} = K_{Li} \times Q_{Li}(t) + K_{Lu} \times Q_{Lu}(t) + K_V \times Q_V(t)$$

The indices B, F, M, Li, Lu, V, and Met stand, respectively, for blood, fat, muscle, liver, lungs, viscera, and metabolites. The symbol I used in the first equation represents the set $\{F, M, Li, V\}$. For each compartment i, Q_i, V_i, and PC_i are, respectively, the BD quantity, the volume, and the tissue over blood partition coefficient. F_i is the blood flow entering i and K_i the metabolic constant. PC_{Air} is the blood over air partition coefficient, K_{Vent} the ventilation pulmonary rate and C_{inh} the inhaled BD concentration. The exhaled BD concentration is given by the following relationship,

$$C_{exh}(t) = \frac{1}{3} \times C_{inh}(t) + \frac{2}{3} \times \frac{Q_{Lu}(t)}{PC_{Lu} \times V_{Lu}} \tag{17}$$

The experimental conditions simulated (an inhalation exposure of 5 ppm during 2 hr) were defined on the basis of actual experiments (18).

A Constrained Lumping Scheme

To apply numerical constrained lumping to the differential equation system (Eq. 15), parameters were set to the physiological values referred in (20). Given these values, the matrix of the system coefficients, **K**, is

$$\mathbf{K} = \begin{pmatrix} -244.90 & 0.02 & 1.16 & 10.99 & 160.20 & 30.11 & 0 \\ 4.41 & -0.02 & 0 & 0 & 0 & 0 & 0 \\ 44.20 & 0 & -1.16 & 0 & 0 & 0 & 0 \\ 19.59 & 0 & 0 & -210.72 & 0 & 0 & 0 \\ 122.45 & 0 & 0 & 0 & -610.60 & 0 & 0 \\ 54.24 & 0 & 0 & 0 & 0 & -43.61 & 0 \\ 0 & 0 & 0 & 199.73 & 13.50 & 13.50 & 0 \end{pmatrix} \tag{18}$$

To construct the lumping matrix **M**, we first need to define the submatrix $\mathbf{M_G}$ containing the constraints. Suppose the state variables are sorted in the same order as in the system (Eq. 16). Given the constraint defined above (quantity metabolized left unlumped), $\mathbf{M_G}$ is

$$\mathbf{M_G} = \begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{19}$$

To determine **M**, we apply the DCAL method presented in Appendix A. Equation 7 states that the Jacobian matrix of the system is **K**. Therefore $\mathbf{A_0}$ is equal to the transpose of **K**. In order to force $\mathbf{M_G}$ to be located on

the first rows of **M**, we multiply $\mathbf{M_G}$ (in Eq. 24, Appendix A) by 1000 (see Appendix A). We then obtain the matrix **Y**,

$$\mathbf{Y}=\begin{pmatrix} 0.342 & -1.19\times10^{-5} & -6.93\times10^{-4} & -0.560 & -0.793 & -2.18\times10^{-2} & 0 \\ -1.19\times10^{-5} & 0 & 0 & 1.56\times10^{-5} & 2.83\times10^{-5} & 0 & 0 \\ -6.93\times10^{-4} & 0 & 0 & 9.09\times10^{-4} & 1.65\times10^{-3} & 4.28\times10^{-5} & 0 \\ -0.560 & 1.56\times10^{-5} & 9.09\times10^{-4} & 2.725 & 1.131 & 0.106 & 0 \\ -0.793 & 2.83\times10^{-5} & 1.65\times10^{-3} & 1.131 & 1.926 & 5.41\times10^{-2} & 0 \\ -2.18\times10^{-2} & 0 & 4.28\times10^{-5} & 0.106 & 5.41\times10^{-2} & 5.94\times10^{-3} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1\times10^{6} \end{pmatrix} \tag{20}$$

Its eigenvalues are (1×10^6, 3.78, 1.21, 0.011, 3.07×10^{-4}, 0, 0), and the corresponding eigenvectors

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1.00 \\ -0.260 & 0 & 4.87\times10^{-4} & 0.769 & 0.583 & 3.16\times10^{-2} & 0 \\ 0.253 & -1.13\times10^{-5} & -6.56\times10^{-4} & 0.637 & -0.728 & 1.90\times10^{-2} & 0 \\ 0.870 & -1.85\times10^{-5} & -1.07\times10^{-3} & 2.58\times10^{-2} & 0.335 & 0.361 & 0 \\ -0.334 & 2.46\times10^{-4} & 1.43\times10^{-2} & -4.91\times10^{-2} & -0.135 & 0.931 & 0 \\ 6.01\times10^{-3} & 1.75\times10^{-2} & 1.00 & 7.74\times10^{-4} & 1.53\times10^{-3} & -1.30\times10^{-2} & 0 \\ 0 & -1.00 & 1.75\times10^{-2} & 0 & 0 & 0 & 0 \end{pmatrix} \tag{21}$$

Values whose magnitude was less than 10^{-5} were set to zero. Two eigenvalues of Y are zero. This implies that aggregation is exact for lumped systems of dimension 6. The lumping matrix of dimension m is constructed with the first m eigenvectors.

Effect of Increased Lumping on the Prediction of the Quantity of BD Metabolized

Dimensions 2–6 were tested for the lumped model. Figure 2 shows the results of constrained lumping of the BD model. The predicted BD quantity metabolized is represented for systems with dimensions 2–6 (100 time-points were used to produce the graph). Predicted quantity of BD metabolized versus time curves can be split into two parts: during exposure, the quantity metabolized increases, and then comes close to an upper bound at the end of inhalation. As expected, the lower the model dimension, the higher the error. Yet, the sign of errors can also change abruptly. About 8 hr after exposure, the solution of the system with dimension 6 does not differ by more than 0.3% from the one of the initial system. For a system of dimension 5, the maximum relative error is 20%. Error goes up to 93% for dimension 3, and to 138% for dimension 2.

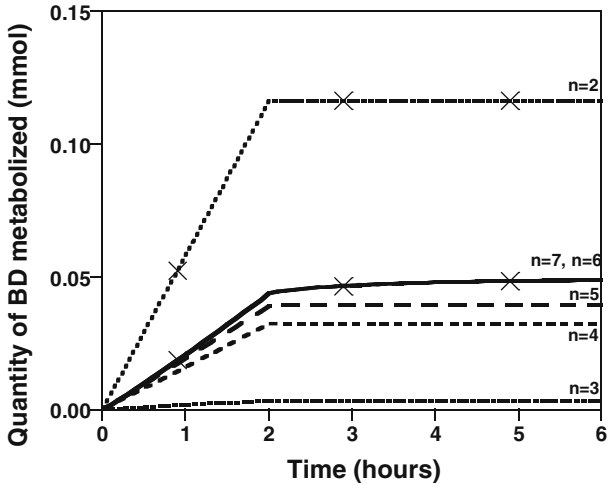

Fig. 2. Time evolution of the quantity of BD metabolized, after an exposure to 5 ppm BD for 2 hr, predicted with the original system (solid curve, $n = 7$) and lumped systems with dimension $n = 2 - 6$. 100 time-points were used to produce these graphs. Lumped systems were defined under the constraint that the total quantity of BD metabolized should be left unlumped. The solid curve with the "**X**" corresponds to the prediction for the system of dimension 6 (which overlays the solution of the original model).

The above results were obtained by using only one set of parameter values. To confirm that dimension reduction is robust to uncertainty in parameter values, different parameter sets were obtained by Monte Carlo simulations. These sets were sampled from normal distribution centered on the value used initially with a coefficient of variation of 10%. For 20 random parameter vectors, the relative error between the initial system and the lumped system was then calculated as a function of time. For the lumped system of dimension 6, the maximum error is 0.5% 8 hr after exposure ($t = 10$). Figure 3 shows the time evolution of the relative error between the prediction of Q_{Met} with original system and the prediction of the lumped system of dimension 5, for 20 parameter vectors. The maximum error lies between 14 and 22%, at $t = 10$ (i.e., the end of the simulations).

Increasing the Constraints to Be Able to Fit the Model

Other constraints can be applied. For example, TK models are often developed to be fitted to experimental data, even PBPK models, as demonstrated by Gelman *et al.* (22). If needed, the lumped model may retain the possibility to be fitted to data. This implies that the model variables for which experimental data are available should not be lumped with

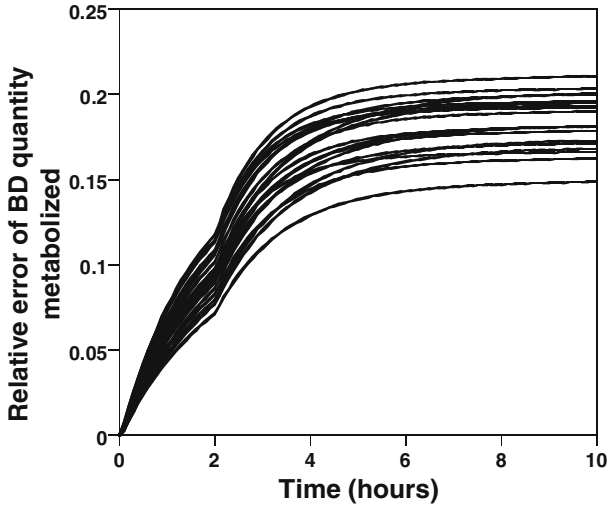

Fig. 3. Time evolution of the relative error (between the prediction of the original system of dimension 7 and the prediction of the lumped system of dimension 5) for the quantity of BD metabolized, after an exposure to 5 ppm for 2 hr. About 20 parameter vectors are shown.

others. In the study conducted by Bois *et al.* (18) (reference for our exposure scenario), exhaled air measurements were performed for each human volunteer. Given Eq. (17), we hence impose that the state variable standing for the BD quantity in lungs, Q_{Lu}, is left unlumped, in addition to the BD quantity metabolized (Q_{Met}). The sub-matrix $\mathbf{M_G}$ containing the constraints is therefore defined by

$$\mathbf{M_G} = \begin{pmatrix} 0\,0\,0\,0\,1\,0\,0 \\ 0\,0\,0\,0\,0\,0\,1 \end{pmatrix} \tag{22}$$

Similar calculations as above were performed. For the lumped system of dimension 6, the predicted quantity metabolized does not differ by more than 0.5% (at $t = 10$) from the prediction of the original system. The deviations of the predicted BD concentration in exhaled air from the original solution are presented in Fig. 4. The relative error increases rapidly with observation time. About 4 hr after the beginning of exposure ($t = 4$), this error is less than 4%. It increases to 10% 1 hr later ($t = 5$), and reaches 90% at $t = 10$. Given these results, how to conclude on the utility of the lumped system of dimension 6? The increase in relative error for C_{exh} is linked to the fact that the concentration of BD in exhaled air decreases rapidly. At the end of exposure ($t = 2$), C_{exh} is equal to 4.6 ppm, at $t = 4$ C_{exh}

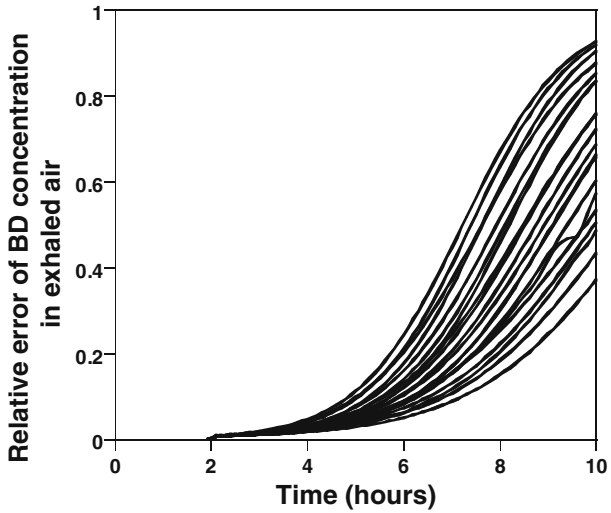

Fig. 4. Time evolution of the relative error (between the prediction of the original system and the prediction of the lumped system of dimension 6) of the concentration of BD in exhaled air, after an exposure to 5 ppm for 2 hr. The total quantity of butadiene metabolized and the quantity in lungs were left unlumped. Twenty parameter vectors are shown.

is 0.023 ppm, at $t = 5$ C_{exh} is 0.010 ppm, and at $t = 10$ C_{exh} is 0.0008 ppm. At $t = 6$, C_{exh} is already under the limit of quantitation, i.e. 0.006 ppm (21). In this example, C_{exh} was left unlumped in order to fit the lumped model. Since experimental data could not be obtained 6 hr after the beginning of exposure, performance of the lumped model after that time is irrelevant. We can then define a criterion which accepts the lumped model if the relative error of its predictions at the limit of quantitation does not exceed a 10th (for example) of the relative error of measurements at the same limit (which is about 200%). At $t = 5$, the relative error induced by the lumped system is 10%, which fulfils the criterion. The lumped model of dimension 6 can therefore be used for our application, and the loss of information (compared to the original system) is negligible.

For the lumped system of dimension 5, we obtained a maximum relative error for Q_{Met} equivalent to the one of the first lumping scheme, that is around 20%, which might be deemed acceptable. However, the prediction of C_{exh} is not at all satisfying according to the above criterion. During exposure (t between 0 and 2), the error is less than 5%. But at t =5, C_{exh} predicted by the lumped system tends rapidly to zero, and the maximum relative error is equal to 100%, which is not acceptable.

DISCUSSION

In this work, we have presented mathematical lumping methods and applied them to TK/PK models. That approach has been initially presented by Wei and Kuo (10), and improved later by Li and Rabitz (6,7,11,15,23,24). These authors proposed a formal method to reduce first-order ordinary differential equation systems, based on a mathematical analysis of the entire system. Such lumping methods are easily automated, and can be used systematically in the search for an optimal lumping. For PK modelers, such an approach to lumping might seem not very intuitive, unlike the one developed by Nestorov *et al.* (4) which takes into account the biochemical properties of the substance (i.e., on the values taken by the model parameters, the matrix **K**). A formal approach can nevertheless be used to support the modeler's intuition. A really convincing result is obtained if the formal procedure points to a lumping scheme similar to an "intuitive" approach. On the other hand, a really interesting result is obtained if the methods disagree strongly (but in that case an understanding of the reason why would be worth seeking). Moreover, it is not always possible to use intuitive methods in case of very large systems, or when we want to find the smallest system dimension needed to describe the PK of a substance in a particular compartment. In such cases, our approach has the potential to reveal similarities in variables that may not be apparent in an analysis solely based on considerations of their attributes (14). In practice, we often need the lumped system to meet predefined (and precise) goals. To that aim, some state variables (e.g. concentrations) should not be lumped with others. Constraints can also follow from the modeler's intuition. For example, one can choose to lump two compartments in which the substance exhibits similar PK profiles. Usually, with such constraints, the system can only be approximately lumped. Constraints should then be used with parsimony, since the accuracy of the lumped model predictions decreases as the number of constraints increases.

In our examples, we have first treated a 2-compartment model by symbolic lumping. Reduced systems obtained by symbolic lumping are valid for all possible parameter values satisfying conditions specified in Eq. (2). Numerical lumping follows the same rules but works with a fully specified original model (including numerical values of the parameter matrix **K**). Calculations can then be done numerically, but the lumped systems obtained are specific to the parameter values used. Obviously, symbolic lumping is more general and flexible than numerical lumping. However the calculation of eigenvalues and eigenvectors cannot usually be performed symbolically for large dimensional matrices/models. Typically, low dimensional models (such as empirical PK models) can be treated symbolically as well as numerically, whereas high

dimensional models (such as PBPK ones) are best treated numerically. However, large systems having special symmetries may be lumped symbolically. A compromise between these two approaches is to set some parameters and lump (if possible) the system with a few symbolic parameters.

Lumping can be appropriate to overcome statistical identifiability problems in parameter estimation. Such problems arise, for example, when model parameters are highly correlated or have multiple peak posteriors distributions (or likelihood functions). In such cases, parameter estimation requires a long time to explore the space of possible parameter values. Work with a system of reduced dimension is therefore be beneficial for such problems. However, in cases of structural identifiability problems (i.e. parameters are not mathematically identifiable), lumping approaches are not appropriate tools.

We also treated the reduction of a 7-compartment PBPK model for 1,3-butadiene (BD) with numerical constrained lumping. For that example, we imposed two constraints on the state variables of the new lumped system: the quantity of BD metabolized (Q_{Met}), and the concentration in exhaled air (C_{exh}) should be left unlumped. Only Q_{Met} was unlumped in Scenario I, and both were unlumped in Scenario II. In these cases, model reduction led to a set of approximate models. The lumped models were tested to check if they gave appropriate answers to actual TK questions. To insure that a lumped model is appropriate, it is necessary to define a quantitative criterion for each value of interest. In the BD example, our objective was to quantify the amount of BD metabolized by humans after exposure. For Scenarios I and II, we show that the original system of dimension 7 can be successfully reduced into a 6-dimensional system (less than 1% of information for Q_{Met} is lost). The reduction into a 5-dimensional system is not really satisfying for both scenarios. We can therefore conclude that all but one compartments in the original system are necessary to correctly describe BD biodistribution, and to quantify reliably its fraction metabolized.

Physiological PK models and other *in silico* tools for fast throughput screening will certainly undergo a development and increase in complexity in the future. While adding complexity to a model is a relatively easy heuristic task, simplifying a complex model is typically more difficult, since that is in itself a complicated problem. Formal lumping techniques, such as the one presented here, have the potential to automatically reduce models and tailor them to specific needs (statistical validation versus predictions, etc.). The process could be made transparent to the user and information gained from the reduced model could be transferred back to the full model via Bayesian updating, for example.

ACKNOWLEDGMENTS

This research was partly supported by the French Ministry of the Environment and Sustainable Development (BCRD-AP2001-DRC07), and by the United States National Institutes of Health grant T32 ES07142-18. Partial support from the National Science Foundation, Hungary (Nos T 037491 and T047132) are acknowledged by János Tóth. The present work is part of the cooperation between INERIS and BUTE supported by the French Ministry for Ecology and Sustainable Development under no. BRCD AP 2001.

APPENDIX A

The DCAL method (14–16) determines the base matrices $\mathbf{A_k}$ numerically by using values of $\mathbf{y}$ in a region of the n-dimensional composition space of interest. It can be shown that the lumping matrix $\mathbf{M}$ can be determined from

$$\mathbf{X}^{\mathrm{T}}\mathbf{M}^{\mathrm{T}} = 0 \tag{23}$$

where $\mathbf{X}$ is "most nearly" orthogonal to

$$\mathbf{Z} = \begin{pmatrix} \mathbf{M_G} \\ \mathbf{M_G}\mathbf{A}_1^{\mathrm{T}} \\ \cdots \\ \mathbf{M_G}\left(\mathbf{A}_1^{\mathrm{T}}\right)^{s_1-1} \\ \cdots \\ \mathbf{M_G} \\ \mathbf{M_G}\mathbf{A}_{\mathrm{m}}^{\mathrm{T}} \\ \cdots \\ \mathbf{M_G}\left(\mathbf{A}_{\mathrm{m}}^{\mathrm{T}}\right)^{s_m-1} \end{pmatrix} \tag{24}$$

where s_k is the rank of $\mathbf{A}_k$. Briefly, the subsequent steps to determine $\mathbf{M}$ are:

- Transform the matrix product $\mathbf{M_G}(\mathbf{A}_k^{\mathrm{T}})^i$ (i from 0 to $s_k - 1$) into an orthogonal matrix $\mathbf{Q}(\mathbf{G})_{ki}^{\mathrm{T}}$ by Gram-Schmidt orthogonalization (25),
- Construct the symmetric matrix $\mathbf{Y}$ defined by:

$$\mathbf{Y} = \sum_{k=1}^{m} \sum_{i=0}^{s_k-1} \mathbf{Q}\,(\mathbf{G})_{k_i}^{T} .\mathbf{Q}\,(\mathbf{G})_{k_i} \tag{25}$$

- Determine the eigenvalues and eigenvectors of $\mathbf{Y}$.

The matrix **X** is constructed with the eigenvectors corresponding to the smallest $n - \hat{n}$ eigenvalues of **Y**. The eigenvectors corresponding to the largest $\hat{n}$ eigenvalues of **Y** comprise the matrix **M**. To force $\mathbf{M_G}$ to correspond to the eigenvectors of **Y** with the largest eigenvalues, $\mathbf{M_G}$ in **Z** is multiplied by a large constant (15). This ensures that the unlumped species specified by $\mathbf{M_G}$ are part of the lumped system. The obtained matrix **M** corresponds to the best constrained approximate lumping matrix with dimension $\hat{n}$, according to the DCAL method. If m eigenvalues of **Y** are equal to zero, the first n–m eigenvectors of **Y** compose an exact lumping matrix $\mathbf{M}^{\mathrm{T}}$.

REFERENCES

1. M. Gibaldi and D. Perrier. Pharmacokinetics. Second Edition, revised and expanded ed., Marcel Dekker, New York, 1982.
2. J. E. Dennison, M. E. Andersen, and R. S. H. Yang. Characterization of the pharmacokinetics of gasoline using PBPK modeling with a complex mixtures chemical lumping approach. *Inhal. Toxicol*. **15**(10):961–986 (2003).
3. J. E. Dennison, M. E. Andersen, I. D. Dobrev, M. M. Mumtaz, and R. S. H. Yang. PBPK modeling of complex hydrocarbon mixtures: Gasoline. *Environ. Toxicol. Pharmacol*. **16**(1–2):107–119 (2004).
4. I. A. Nestorov, L. J. Aarons, P. A. Arundel, and M. Rowland. Lumping of whole-body physiologically based pharmacokinetic models. *J. Pharmacokinet. Biopharm*. **26**:21–46 (1998).
5. S. Bjorkman. Reduction and lumping of physiologically based pharmacokinetic models: Prediction of the disposition of fentanyl and pethidine in humans by successively simplified models. *J. Pharmacokinet. Pharmacodyn*. **30**(4):285–307 (2003).
6. G. Li and H. Rabitz. A general analysis of exact lumping in chemical kinetics. *Chem. Eng. Sci*. **44**(6):1413–1430 (1989).
7. G. Li and H. Rabitz. A general analysis of approximate lumping in chemical kinetics. *Chem. Eng. Sci*. **45**(4):977–1002 (1990).
8. G. Y. Li, A. S. Tomlin, H. Rabitz, and J. Toth. A general analysis of approximate nonlinear lumping in chemical kinetics. II. Unconstrained lumping. *J. Chem. Phy*. **101**(2):1172–1187 (1994).
9. G. Li, H. Rabitz, and J. Toth. A general analysis of exact nonlinear lumping in chemical kinetics. *Chem. Eng. Sci*. **49**(3):343–361 (1994).
10. J. Wei and J. C. W. Kuo. A lumping analysis in monomolecular reaction systems. Analysis of the exactly lumpable system. *Ind. Eng. Chem. Fundam*. **8**:114–123 (1969).
11. G. Li and H. Rabitz. The direct lumping approach: An application to a catalytic reforming model. *Chem. Eng. Sci*. **48**(10):1903–1909 (1993).
12. A. S. Tomlin, T. Turanyi, and M. J. Pilling, *Mathematical Tools for the Construction, Investigation and Reduction of Combustion Mechanisms*. In *Low-temperature Combustion and Autoignition*, M. J. Pilling (ed.) Elsevier, Amsterdam, pp. 293–437 (1997).
13. H. J. Verhaar, J. R. Morroni, K. F. Reardon, S. M. Hays, D. P. Gaver Jr., R. L. Carpenter, and R. S. Yang. A proposed approach to study the toxicology of complex mixtures of petroleum products: The integrated use of QSAR, lumping analysis and PBPK/PD modeling. *Environ. Health. Perspect*. **105** (Suppl 1):179–195 (1997).
14. S. W. Wang, P. G. Georgopoulos, G. Li, and H. Rabitz. Condensing complex atmospheric chemistry mechanisms. 1. The direct constrained approximate lumping (DCAL) method applied to alkane photochemistry. *Environ. Sci. Technol*. **32**(13): 2018–2024 (1998).

15. G. Li and H. Rabitz. New approaches to determination of constrained lumping schemes for a reaction system in the whole composition space. *Chem. Eng. Sci.* **46**(1):95–111 (1991).
16. G. Li and H. Rabitz. Determination of constrained lumping schemes for nonisothermal first-order reaction systems. *Chem. Eng. Sci.* **46**(2):583–596 (1991).
17. J. A. Jacquez and C. P. Simon. Qualitative theory of compartmental systems. *Siam Rev*. **35**:43–79 (1993).
18. F. Y. Bois, T. J. Smith, A. Gelman, H. Y. Chang, and A. E. Smith. Optimal design for a study of butadiene toxicokinetics in humans. *Toxicol. Sci*. **49**(2):213–224 (1999).
19. M. Mezzetti, J. G. Ibrahim, F. Y. Bois, L. M. Ryan, L. Ngo, and T. J. Smith. A Bayesian compartmental model for the evaluation of 1,3-butadiene metabolism. *J. Roy. Stat. Soc. C-App. Stat.* **52**:291–305 (2003).
20. M. C. Kohn and R. L. Melnick. Species differences in the production and clearance of 1,3-butadiene metabolites: a mechanistic model indicates predominantly physiological, not biochemical, control. *Carcinogenesis* **14**:619–628 (1993).
21. T. Smith, Y.-S. Lin, L. Mezzetti, F. Y. Bois, K. Kelsey, and J. Ibrahim. Genetic and dietary factors affecting human metabolism of 1,3-butadiene. *Chem. Biol. Interact*. 135–136:407–428 (2001).
22. A. Gelman, F. Y. Bois, and J. Jiang. Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *J. Am. Stat. Assoc*. **91**(436):1400–1412 (1996).
23. G. Li and H. Rabitz. A general lumping analysis of a reaction system coupled with diffusion. *Chem. Eng. Sci*. **46**(8):2041–2053 (1991).
24. G. Li and H. Rabitz. Reduced kinetic equations of a CO/H2/air oxidation model by a special perturbation method. *Chem. Eng. Sci*. **52**(23):4317–4327 (1997).
25. R. Bellman. Introduction to Matrix Analysis. McGraw-Hill Book Company Inc. New York, Toronto, London (1960).