*J. Inv. Ill-Posed Problems*, Vol. 13, No. 1, pp. 1–26 (2005)


# A simulation-based comparison between parametric and nonparametric estimation methods in PBPK models

H. T. BANKS*, Y. MA*, and L. K. POTTER†

*Received June 29, 2004*

**Abstract** — We compare parametric and nonparametric estimation methods in the context of PBPK modeling using simulation studies. We implement a Monte Carlo Markov Chain simulation technique in the parametric method, and a functional analytical approach to estimate the probability distribution function directly in the nonparametric method. The simulation results suggest an advantage for the parametric method when the underlying model can capture the true population distribution. On the other hand, our calculations demonstrate some advantages for a nonparametric approach since it is a more cautious (and hence safer) way to assess the distribution when one does not have sufficient knowledge to assume a population distribution form or parametrization. The parametric approach has obvious advantages when one has significant *a priori* information on the distributions sought, although when used in the nonparametric method, prior information can also significantly facilitate estimation.

## 1. INTRODUCTION

In this paper we compare estimation procedures using random effects type techniques with those using a Bayesian based Monte Carlo Markov Chain (MCMC) approach. In the random effects type approach we follow the Prohorov Metric Framework (PMF) formulation as developed by Banks, Bihari and Fitzpatrick in several papers [2, 3, 6]. Early versions of these ideas were developed in the

*Center for Research in Scientific Computation, North Carolina State University, Raleigh, N.C. 27695-8205. E-mails: htbanks@ncsu.edu, yma@ncsu.edu

†Scientific Computing and Mathematical Modeling, GlaxoSmithKline, Research Triangle Park, NC 27709. E-mail: laura.k.potter@gsk.com

This research was supported in part by the Joint DMS/NIGMS Initiative to Support Research in the Area of Mathematical Biology (grant 1R01GM67299-01), in part by the US Air Force Office of Scientific Research (grant FA9550-04-1-0220), and by the Statistical and Applied Mathematical Sciences Institute (SAMSI), which is funded by National Science Foundation (grant DMS-0112069).

context of specific size structured partial differential equation population models [4, 5, 10]. These efforts motivated a formulation in a more general functional analytic framework for estimation of underlying (absolutely) continuous distributions or measures. The approach employs approximations with finite convex combinations of Dirac measures which can be guaranteed to converge as the finite number increases. This PMF formulation is related to, but distinctively different from, the mixing distribution or nonparametric maximum likelihood (NPML) formulations of Lindsay [18, 19] and Mallet [12, 20, 26] wherein one uses convex geometry (Caratheordory based representations) to justify seeking a fixed dimension (the number of underlying distinct likelihoods used in the maximum likelihood estimator or MLE process) convex combination of Dirac measures as a (generally non-unique) MLE.

Here we use the PMF formulation in the context of ordinary least squares (OLS) estimators, primarily to facilitate exposition. An MLE formulation would produce quite similar results with respect to both efficiency of approximation and computational times in comparison to MCMC techniques.

We compare these methods on a typical physiologically based pharmacokinetic (PBPK) model (chosen simply to illustrate the method behaviors) that is detailed in Section 2. In Section 3 we outline the MCMC/Gibbs-Metropolis-Hastings algorithm that we use as it is embodied in the software package MCSim. The proposed PMF method is given in Section 4 along with a brief overview of the Prohorov metric and convergence results. Finally, sample results from 78 examples (based on a significant computational effort reported in detail in [7]) are provided in Section 5 along with some summary conclusions on our findings in a final section.

## 2. THE PBPK MODEL

In this section we provide an overview of the PBPK model for TCE as developed in [1, 25]. This model utilizes standard physiologically based pharmacokinetic compartmental equations that are based on assumptions of rapid well-mixing and equilibrium.

Many PBPK models for lipophilic compounds include compartments for tissues such as the liver, lungs, adipose tissue, richly perfused and poorly perfused tissues. These compartments often assume a perfusion-limited model, or equivalently, a flow-limited model of disposition, meaning that the rate of uptake of the compound into the tissue is limited by the blood flow rate to the tissue rather than the rate of diffusion across the cell membranes [21]. In this case, the blood flow rate to the tissue is slow compared to the diffusion rate across cell membranes, so that the blood and tissue are in equilibrium. The equation for transport of a solute through a constant-volume, well-mixed tissue compartment is an ordinary differential equation of the form

$$V \frac{dC}{dt} = Q(C_{in} - C_{out}),$$

where $V$ is the volume of the tissue (in liters), $C$ is the concentration of com-

pound inside the tissue (in mg/liter), $Q$ is the blood flow rate to the tissue (in liters/hour), and $C_{in}$ and $C_{out}$ are the compound concentrations entering and exiting the tissue, respectively.

Here we present a standard PBPK model [21] for TCE with flow-limited compartments for the kidney, muscle tissue, adipose tissue, brain, liver, venous blood, and remaining non-fat tissue (see Figure 1). As detailed in [25], we assume uptake via inhalation, with a lung compartment subdivided into the alveolar space and lung blood subcompartments. TCE is metabolized in the liver, which is modeled with Michaelis-Menten kinetics.

In the lung, ventilation is assumed to be continuous with rate $Q_p$, and the vapor in the alveolar space is assumed to be in rapid equilibrium with the arterial lung blood. The cardiac output rate is given by $Q_c$ and the blood/air partition coefficient is denoted by $P_b$.

The variables used in the lung compartment include:

$$\begin{aligned} C_c &= \text{Concentration of TCE in surrounding air} \\ C_x &= \text{Concentration of TCE in alveolar space} \\ C_a &= \text{Concentration of TCE in arterial blood} \\ C_v &= \text{Concentration of TCE in venous blood} \\ A_i &= \text{Amount of TCE inhaled} \\ A_x &= \text{Amount of TCE exhaled} \\ A_L &= \text{Amount of TCE in lung.} \end{aligned}$$

In this case, the concentration $C_x$ in the alveolar air is related linearly to the concentration $C_a$ in the arterial blood:

$$C_x = \frac{C_a}{P_b}.$$

The rate of inhalation of TCE is given by $Q_p C_c$, while the rate of exhalation is given by $Q_p C_x$. Therefore we have the following equations:

$$\begin{aligned} \frac{dA_i}{dt} &= Q_p C_c, \\ \frac{dA_x}{dt} &= Q_p C_x = Q_p \frac{C_a}{P_b}, \\ \frac{dA_L}{dt} &= Q_p(C_c - C_x) + Q_c(C_v - C_a). \end{aligned} \tag{2.1}$$

Moreover, the assumptions of the model [21] imply that $dA_L/dt = 0$, so by substituting $C_x = C_a/P_b$ into (2.1) we obtain

$$C_a = \frac{Q_p C_c + Q_c C_v}{Q_c + Q_p/P_b}.$$

Combining the perfusion-limited compartments with the lung compartment, we obtain

$$V_f \frac{dC_f}{dt} = Q_f(C_a - C_{vf}), \tag{2.2}$$

$$V_v \frac{dC_v}{dt} = Q_m C_{vm} + Q_t C_{vt} + Q_f C_{vf} \tag{2.3}$$

$$+ Q_{br} C_{vbr} + Q_l C_{vl} + Q_k C_{vk} - Q_c C_v, \tag{2.4}$$

$$C_a = \frac{Q_c C_v + Q_p C_c}{Q_c + Q_p/P_b}, \tag{2.5}$$

$$V_m \frac{dC_m}{dt} = Q_m(C_a - C_{vm}), \tag{2.6}$$

$$V_t \frac{dC_t}{dt} = Q_t(C_a - C_{vt}), \tag{2.7}$$

$$V_{br} \frac{dC_{br}}{dt} = Q_{br}(C_a - C_{vbr}), \tag{2.8}$$

$$\frac{dA_{am}}{dt} = \frac{v_{\max} C_{vl}}{k_M + C_{vl}}, \tag{2.9}$$

$$V_l \frac{dC_l}{dt} = Q_l(C_a - C_{vl}) - \frac{v_{\max} C_{vl}}{k_M + C_{vl}}, \tag{2.10}$$

$$V_k \frac{dC_k}{dt} = Q_k(C_a - C_{vk}), \tag{2.11}$$

$$\frac{dA_i}{dt} = Q_p C_c, \tag{2.12}$$

$$\frac{dA_x}{dt} = Q_p C_x. \tag{2.13}$$

The subscripts denote the following specific tissues:

$$\begin{aligned} v &\Leftrightarrow \text{Venous blood},\\ k &\Leftrightarrow \text{Kidney},\\ m &\Leftrightarrow \text{Muscle},\\ f &\Leftrightarrow \text{Fat},\\ br &\Leftrightarrow \text{Brain},\\ l &\Leftrightarrow \text{Liver},\\ t &\Leftrightarrow \text{Remaining non-fat tissue}. \end{aligned}$$

Volumes (in liters) of specific tissues are denoted by $V$, concentrations of TCE (mg/liter) are denoted by $C$ and flow rates (liters/hour) are denoted by $Q$, each with subscripts corresponding to the specific tissue. The concentration of TCE in the air is denoted by $C_c$, and is a specified quantity. The variables $C_{vk}, C_{vm}, C_{vf}, C_{vbr}, C_{vl}$ and $C_{vt}$ are the concentrations of TCE leaving the respective organ and entering the venous blood system. In this case, all compartments except for the lung are perfusion-limited, so the concentration of

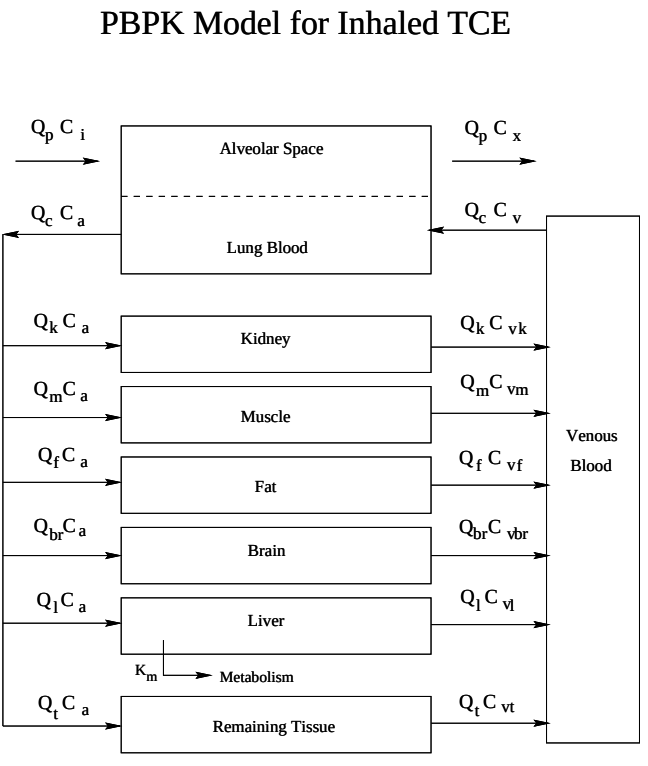


Figure 1. Schematic of PBPK model for inhaled TCE in Long-Evans rats

TCE leaving each of these compartments is equal to the concentration of free TCE in that compartment itself [21]. In the kidney, for example, this implies

$$C_{vk} = \frac{C_k}{P_k},$$

where $C_k$ is the total concentration of TCE inside the kidney compartment and $P_k$ is the tissue/blood partition coefficient for the kidney.

The amount of TCE metabolized in the liver is denoted by $A_{am}$, and has units in milligrams. Constants in the liver compartmental model include the Michaelis-Menten constant $k_M$ (mg/liters) and the metabolic constant $v_{\max}$ (mg/hour).

The parameters in the system may be treated as realizations of random variables, especially when using the model with data from a population of individuals. We refer to the system given by equation through (2.1) to (2.13) as the **mathematical model**. Since the model describes concentrations of TCE within the organs and tissues of an individual, the parameter set in the mathematical model is individual specific. In the following context, we use

bold face letters to represent vectors. We assume we have observations $\boldsymbol{x}_{ij}$, $j = 1, \ldots, n_i$, measured at (possibly different) times $t_{ij}$, $j = 1, \ldots, n_i$, for individuals $i$, $i = 1, \ldots, n$. We represent our data as $D = \{(t_{ij}, \boldsymbol{x}_{ij}) : i = 1, \ldots, n, j = 1, \ldots, n_i\}$. We further assume our data are subject to normally distributed error. Here we generate synthetic data using the mathematical model and then add simulated measurement error. The data are centered at the solution of the mathematical model with parameters $\boldsymbol{q}_i$, and have constant variance. In other words, we assume

$$\boldsymbol{x}_{ij} = \boldsymbol{y}(t_{ij}, \boldsymbol{q}_i) + \boldsymbol{\epsilon}_{ij}, \qquad \boldsymbol{\epsilon}_{ij} \sim N(0, \Sigma_\epsilon), \tag{2.14}$$

where $\boldsymbol{y}(\cdot, \boldsymbol{q}_i)$ represents the solution to the system given in the math model for the $i$th individual and $\Sigma_\epsilon$ is the covariance matrix of the distribution of $\boldsymbol{\epsilon}_{ij}$. The assumption we make through equation (2.14) is part of a **statistical model**.

Note that the quantity of interest here is the distribution of the $\boldsymbol{q}_i$'s. Depending on the assumptions made about this distribution, one arrives at either a "parametric approach" or a "nonparametric approach" as described in the next two sections.

## 3. THE PARAMETRIC APPROACH

Both mathematicians and statisticians use the term "parametric" when describing approaches to estimation problems. When both the mathematical model and the statistical model are fully specified at all levels, the approach is referred to by statisticians as a parametric approach. Specifically, in the PBPK model, when one assumes a distribution for each individual's measurement, and also a distribution for the set of individual parameters across a population, one is using a parametric approach. On the other hand, mathematicians refer to the approach as parametric when a fully parametrized distribution assumption is made on the parameters to be estimated. This is independent of the assumption (often omitted or only tacitly made in mathematical treatments) of a distribution for measurement error. Moreover, an implicit assumption of the errors being identically distributed with a normal distribution having mean zero and constant variance is tacitly made when the standard approach of ordinary least squares is taken and assumed equivalent to an MLE. Of course, one can use the OLS without the normality assumption (and then it will not be equivalent to the MLE).

A parametric method is appropriate when one is reasonably confident about the general parameter distribution structure in the problem. In such situations, this provides the most efficient way for estimation in that it takes full advantage of knowledge of the distribution structure. However, when a pre-assumed distribution is incorrect, this often leads to serious difficulties. In the best case scenario, the final estimate produces a model that fits the data very poorly. In a more serious outcome, the assumed distribution may produce a reasonable model fit to data even when it is an incorrect distribution (see [9, 25] for examples).

Once a model is specified, one may rely on several different methods to solve the problem and estimate the desired parameters. One of the most widely used methods is the MCMC method [14]. MCMC techniques produce parameter estimates through generation of random samples of the sought after distribution. They are typically used in cases where the distribution does not have a closed form; this often occurs for posterior distributions when a Bayesian approach is taken.

We explain the MCMC approach through a simple example. In this example, we assume for ease in explanation that the parameter of interest $\boldsymbol{q}$ is, in fact, a scalar which we denote by $q$. Moreover we measure only one quantity, for example, one component of the vector solution of (2.1)–(2.13) above. Hence, the measurement $\boldsymbol{x}$ is also a scalar, which we denote by $x$. The measurement error $\boldsymbol{\epsilon}$ is thus a scalar and will be denoted by $\epsilon$. Suppose we have $k$ observations $(t_1, x_1), (t_2, x_2), \ldots, (t_k, x_k)$ associated with $y(t_1, q), \ldots, y(t_k, q)$ and measurement errors $\epsilon_1, \ldots, \epsilon_k$ respectively, on a single individual with dynamics given by (2.1)–(2.13). We then have

$$x_j = y(t_j, q) + \epsilon_j, \qquad j = 1, 2, \ldots, k,$$

where the $\epsilon_j$'s are assumed independent and to follow a normal distribution $N(0, \sigma^2)$. We wish to estimate $q$ and $\sigma^2$ using this data.

The Bayesian approach involves considering $q$ and $\sigma^2$ as random variables. Since we know very little *a priori* about $q$ and $\sigma^2$, we can reasonably collect our knowledge about them in a very flat distribution, which is called a "prior distribution". After we collect and use the data $D = \{(t_j, x_j),\, j = 1, \ldots, k)\}$ to improve our knowledge, we would expect to obtain a "posterior distribution" which would (hopefully) reveal more information about the two parameters. We formalize this idea in the equation

$$\begin{aligned} \pi(q, \sigma^2 | D) &= \frac{\pi(q, \sigma^2) L(q, \sigma^2 \mid D)}{\int_\Omega \pi(q, \sigma^2) L(q, \sigma^2 \mid D) dq\, d\sigma^2} \\ &\propto \pi(q, \sigma^2) \Pi_{j=1}^k \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big( - \frac{(x_j - y(t_j, q))^2}{2\sigma^2} \Big). \end{aligned} \tag{3.1}$$

Here, $\pi(q, \sigma^2)$ is the prior, $\pi(q, \sigma^2 \mid D)$ is the posterior, $L$ is the likelihood, and $\Omega$ is the set of possible parameter values.

If, for example, we assume a prior of $\pi(q, \sigma^2) = \exp(-q^2)/\sigma^2$, we obtain

$$\pi(q, \sigma^2 \mid D) \propto \frac{\exp(-q^2)}{(\sigma^2)^{(n+2)/2}} \exp\Big( - \frac{1}{2\sigma^2} \sum_{j=1}^{k} (x_j - y(t_j, q))^2 \Big). \tag{3.2}$$

While equation (3.2) is not readily familiar as a joint density function of $q$ and $\sigma^2$, the conditional density functions of $q$ and $\sigma^2$ may be familiar in special cases. For example, if $y(t, q) = tq$, they are respectively the normal and inverse Gamma distributions. In such a fortunate situation, we can generate the corresponding random samples of a joint distribution by generating random samples

of the conditional distribution for each random variable. This technique is called Gibbs sampling [14].

The steps in the algorithm are:

1. Set $i = 0$ and give an initial guess for a sample $(q^0, (\sigma^2)^0)$.

2. Generate a random sample $(\sigma^2)^{i+1}$ from the conditional distribution of $\sigma^2$ conditioned on the last generated value $q^i$ of $q$.

3. Generate a random sample $q^{i+1}$ from the conditional distribution of $q$ conditioned on the last generated value $(\sigma^2)^{i+1}$ of $\sigma^2$.

4. If convergence is obtained, stop. Otherwise, set $i = i + 1$ and return to step 2.

In the general case of a vector parameter $\boldsymbol{q} = (q_1, \ldots, q_m)$, we must estimate the joint distribution of $(q_1, q_2, \ldots, q_m)$. The algorithm above is then carried out component-wise. That is, step 3 above is replaced by

3′. Draw a sample $q_1^{i+1}$ from the distribution for $q_1$ conditioned on $q_2^i, \ldots, q_m^i$.
Draw a sample $q_2^{i+1}$ from the distribution for $q_2$ conditioned on $q_1^{i+1}, q_3^i, \ldots, q_m^i$.
Draw a sample $q_3^{i+1}$ from the distribution for $q_3$ conditioned on $q_1^{i+1}, q_2^{i+1}, q_4^i, \ldots, q_m^i$.
.
.
.
Draw a sample $q_m^{i+1}$ from the distribution for $q_m$ conditioned on $q_1^{i+1}, \ldots, q_{m-1}^{i+1}$.

The resulting new draw of the joint distribution is then $(q_1^{i+1}, q_2^{i+1}, \ldots, q_m^{i+1})$.

A discussion on convergence issues for the Gibbs algorithm is given in [11]. In general, many investigators discard the first $10^3$ to $10^4$ samples (this is called the "burn-in" cycle and practice varies widely on the number of discards), retaining only the last sample to be used as an "initial" sample for a larger number of steps of the algorithm. It is then expected that the resulting samples will be very much like the "real" samples following the joint distribution.

If at least one of the conditional distributions is not of a form from which one knows how to generate samples, for example $y(t, q) = e^{qt}$ in the above example, one usually applies the Metropolis-Hastings sampling algorithm [22, 15], which we briefly describe next.

Denote by $\pi(v)$ the density function from which we wish to sample (in the example above, $v = (q, \sigma^2)$). To carry out the Metropolis-Hastings algorithm, we first must propose a parametrized density function of the same random variable $v$, say $p(v, \gamma)$, where here $\gamma$ is a parameter. In each step of the Metropolis-Hastings algorithm, we generate samples from $p(v, \gamma)$, and accept it as a new sample with a certain probability.

The steps are:

1. Set $i = 0$ and give an initial sample guess $v^0$.
2. Generate a random sample $\tilde{v}$ from $p(v, v^i)$.
3. Calculate $r = \min\left(\dfrac{\pi(\tilde{v})p(v^i, \tilde{v})}{\pi(v^i)p(\tilde{v}, v^i)}, 1\right)$.
4. Generate a random sample $u$ from a uniform distribution $U(0, 1)$.
5. Accept $\tilde{v}$ as $v^{i+1}$ if $u \leq r$, or
   Accept $v^i$ as $v^{i+1}$ otherwise.
6. If convergence is obtained, stop. Otherwise, set $i = i + 1$ and go to step 2.

Convergence of the Metropolis-Hastings algorithm is discussed in [27] using the theory of irreducible Markov chains [24]. In practice, one often chooses the proposal density $p$ to be normal with mean $v^i$ to obtain $v^{i+1}$.

In our example above, the conditional distribution of $\sigma^2$ is still an inverse Gamma distribution independent of the form of $y$. Thus we can still use Gibbs sampling to generate samples for $\sigma^2$, while each time we sample from the conditional distribution of $q$, we may apply the Metropolis-Hastings algorithm. Such a technique of embedding the Metropolis–Hastings sampling algorithm in each step of Gibbs sampling is frequently used in practice and is called the hybrid Gibbs–Metropolis–Hastings algorithm [14].

We note that the $v$ in the Metropolis–Hastings algorithm as well as the $q$ in the Gibbs sampling algorithm can be vectors.

The MCMC approach provides one with a way to deal with distributions that are very complex in form. It facilitates some estimations that are otherwise computationally impossible.

We return finally to the PBPK model of Section 2 that is the focus of our efforts in this paper. In order to pursue a parametric approach, one needs to make further assumptions on the distribution of the $q$'s in the population from which we sample. For this, we assume that for a given sample $\{q_1, \ldots, q_n\}$ from $n$ individuals, we have

$$q_i \sim N(\mu, \Sigma). \tag{3.3}$$

This yields a so-called hierarchical model, in that we have a model at the individual level (given by (2.1)–(2.13)) and one at the population level (given by (3.3)). MCMC methods on hierarchical models have been explored by numerous investigators, see for example [17, 28]. Under the full mathematical and statistical models given through equations (2.1)–(2.13), (2.14) and (3.3), we can derive the posterior distribution of the parameters $\mu, \Sigma_\epsilon, \Sigma$ given by

$$\pi(\mu, \Sigma, \Sigma_\epsilon, q_1, \ldots, q_n \mid D) \propto \pi(\mu, \Sigma_\epsilon, \Sigma)\Pi_{i=1}^n \Pi_{j=1}^{n_i} p(x_{ij} \mid q_i, \Sigma_\epsilon)\Pi_{i=1}^n p(q_i \mid \mu, \Sigma).$$

Note that in this case we cannot write out the explicit form of $p(x_{ij} \mid q_i, \Sigma_\epsilon)$ because the mean $y(t_{ij}, q_i)$ of the distribution is given implicitly through the mathematical model in equations (2.1)–(2.13). However, the Metropolis–Hastings algorithm only requires the ability to evaluate each $p(x_{ij} \mid q_i, \Sigma_\epsilon)$ and hence each $y(t_{ij}, q_i)$ for a proposed $q_i$. Therefore, we only need to solve the differential equation for $y$ corresponding to the $q_i$'s in each Metropolis–Hastings step. Using MCMC in problems where the mean function is implicitly given has been explored in the inverse problem literature, see for example [16, 23].

The MCMC software MCSim [13] is particularly designed for such problems. To our knowledge, MCSim is the only software currently available that handles implicit mean functions given through ODEs. This is the software we use in our simulation study for the parametric approach.

## 4. A NONPARAMETRIC APPROACH IN THE PMF

The Prohorov Metric Framework (PMF) approach focuses on estimating the distribution for the parameters $q$ directly from the data $D = \{(t_{ij}, x_{ij})\}$ without making the *a priori* assumption (3.3). One could again use an MLE formulation. However, with the assumptions we make on the measurement errors $\epsilon_{ij}$ in (2.14), i. e., independent, identically distributed normal, one can equivalently use an OLS formulation as we do here.

We assume as before that the model parameters $q$ are realizations of a random variable with population probability distribution $P$, where $P$ belongs to some probability space $\mathcal{Q}$ that may be infinite dimensional. We define the set $\mathcal{P}(Q)$ of all probability distributions on an admissible parameter space $Q$ and seek a probability distribution function $P^*$ that minimizes the objective function

$$J(P, D) = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n_i} |E[y(t_{ij}, q) \mid P] - x_{ij}|^2 \tag{4.1}$$

over $\mathcal{Q} \subset \mathcal{P}(Q)$, where the expected values are given by

$$E[y(t_{ij}, q) \mid P] = \int_Q y(t_{ij}, q)\, dP(q). \tag{4.2}$$

For simplicity we often choose $\mathcal{Q} = \mathcal{P}(Q)$, but this is not essential and one may readily restrict the family of admissible distributions in certain formulations.

Depending on the choice of the set $\mathcal{Q} \subset \mathcal{P}(Q)$ of probability distributions, this method may be implemented with pre-determined "prior" probability distributions (as with the Monte Carlo method), or it may be used without the pre-specification of a particular probability distribution. For the case when there is a reasonable expectation that a parameter varies across the population in a manner similar to a given probability distribution, the set $\mathcal{Q}$ can be chosen as the space of those parametrized distribution functions (e. g., $\tilde{q} = (\mu, \Sigma)$, the normal distribution $p(q \mid \mu, \Sigma)$ of (3.3)) defined over admissible parameter sets $\tilde{Q}$. For this type of formulation, the distribution functions are uniquely

determined by their parameterizations $\tilde{q}$ (e. g., mean and variance), and hence may be estimated by minimizing (4.1) where now the expectations in (4.2) are given by

$$E[y(t_{ij}, q) \mid P] = \int_Q y(t_{ij}, q) p(q \mid \mu, \Sigma)\, dq,$$

and the minimization in (4.1) is now over all admissible $(\mu, \Sigma)$ in some specified parameter space $\tilde{Q}$.

If it is not possible to predict the expected form of the probability distributions *a priori*, this method also may be used without the specification of prior distributions. In this case, $\mathcal{Q} = \mathcal{P}(Q)$ may be chosen as the space of all probability distributions defined on $Q$. For computational purposes, the estimation problem may then be implemented using finite dimensional approximations to the original infinite dimensional problem. First we define the infinite dimensional set

$$\mathcal{P}_0(Q) \equiv \Big\{ P \in \mathcal{P}(Q) : P = \sum_{k=1}^{l} p_k \Delta_{q_k}, \quad l \in \mathbb{N}^+, \quad q_k \in Q_0, \quad p_k \geq 0, \quad \sum_{k=1}^{l} p_k = 1 \Big\}, \tag{4.3}$$

where $Q_0 = \{q_k\}_{k=1}^{\infty}$ is a given countable, dense subset of the parameter space $Q$ and $\Delta_{q_k}$ is the Dirac delta distribution with atom at $q_k \in Q$. In other words, $\mathcal{P}_0(Q)$ is the set of probability distributions on $Q$ that have finite support in $Q_0$. We then define the finite dimensional set $\mathcal{P}^M = \big\{ P_M \in \mathcal{P}_0(Q) : P_M = \sum_{k=0}^{M} p_k \Delta_{q_k} \big\}$, which we use to define a family of finite dimensional approximation problems. That is, for fixed $\{q_0, q_1, \ldots, q_M\}$ in $Q_0$ with $P_M = \sum_{k=0}^{M} p_k \Delta_{q_k} \in \mathcal{P}^M$, we minimize the objective function

$$\begin{aligned} J(P_M, D) &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n_i} |E[y(t_{ij}, q) \mid P_M] - x_{ij}|^2 \\ &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \Big| \sum_{k=0}^{M} y(t_{ij}, q_k) p_k - x_{ij} \Big|^2 \end{aligned} \tag{4.4}$$

over the space $\mathcal{P}^M$. These precise definitions lead to a well-posed estimation problem, as we shall discuss below. Note that the problem of minimizing the objective function (4.4) corresponds to solving a constrained quadratic programming problem for $\boldsymbol{p} = (p_0, p_1, \ldots, p_M)$ with the constraints $p_k \geq 0$, $\sum_{k=0}^{M} p_k = 1$. There currently exist a number of acceptable computational methods to efficiently solve such finite dimensional approximating problems. Using the Prohorov metric and well-known results from probability theory, one can establish a theoretical framework (including well-posedness, convergence of approximations and method stability) for these probability-based parameter estimation problems.

The Prohorov metric $\rho$ is defined on the space $\mathcal{P}(Q)$ of probability measures on the Borel subsets of $Q$, where $Q$ is a complete metric space with metric $d$. The definition of $\rho$ is not very intuitive and will not be given here (see instead [3]). Rather we point out that convergence in the Prohorov metric is equivalent to weak convergence of measures or distributions (*not densities*). That is, $\rho(P_k, P) \to 0$ is equivalent to $\int_Q f(q)\, dP_k(q) \to \int_Q f(q)\, dP(q)$ for all bounded and uniformly continuous functions $f : Q \to \mathbb{R}$. It is also well known that the metric space $(\mathcal{P}(Q), \rho)$ is complete, and furthermore, $(\mathcal{P}(Q), \rho)$ is compact for all compact sets $Q$.

Banks and Bihari [3] addressed theoretical issues related to estimation problems involving unknown measures. Employing the Prohorov metric, they studied convergence properties of sequences of probability distributions in $\mathcal{P}(Q)$. These results were then applied to a sequence of minimizers for finite dimensional approximations to the estimation problem for (4.4). Here we summarize their findings as they relate to the inverse problems of interest in this paper.

As discussed in [3], it follows that if the mapping $q \to y(t_{ij}, q)$ is continuous, then the convergence $\rho(P_k, P) \to 0$ in the Prohorov metric is equivalent to $E[y(t_{ij}, q) \,|\, P_k] \to E[y(t_{ij}, q) \,|\, P]$, and hence the map $P \to J(P, D)$ of (4.1) is continuous in the $\rho$ topology. Moreover, if the space $Q$ is compact, we have that $(\mathcal{P}(Q), \rho)$ is a compact metric space, which along with the continuity of the map $P \to J(P, D)$ guarantees the existence of a minimizer over $\mathcal{P}(Q)$ for the estimation problem associated with (4.1).

In addition to establishing the existence of a solution for the inverse problem for (4.1), Banks and Bihari developed results related to method stability for this problem. Using finite dimensional approximation techniques, they show in Theorem 4.1 of [3] that the solutions for minimizing (4.1) depend continuously on the data (see [3] for a complete discussion). Moreover, any sequence of minimizers of the finite dimensional problems for (4.4) converge in the Prohorov metric to a minimizer for the original infinite dimensional problem for (4.1). This theorem makes use of the result that the set $\mathcal{P}_0(Q)$ as in (4.3) is dense in the space $\mathcal{P}(Q)$ with respect to the Prohorov metric $\rho$.

In demonstrating the convergence of solutions (we note that PMF convergence guarantees convergence in the distributions; the densities may not converge at all-see [2, 3, 5, 9, 10]) for the family of finite dimensional problems for (4.4), the result established in Theorem 4.1 of [3] also provides a computational framework for solving the general parameter estimation problem for (4.1) without specifying prior probability distributions. Using discrete Dirac delta measures, for sufficiently large $M$ we may approximate

$$\int_Q y(t, q)\, dP(q) \approx \int_Q \sum_{k=0}^{M} y(t, q) d\Delta_{q_k}(q) = \sum_{k=0}^{M} y(t, q_k) p_k,$$

which then allows us to approximate the infinite dimensional inverse problem for (4.1) by the finite dimensional approximation. Alternative approximations in terms of splines (piecewise linear for (4.4), as well as higher order splines) can also be used as explained in [8].

## 5. SIMULATION STUDY

In this section we present results for the nonparametric and parametric parameter estimation methods described in Section 3 and 4 when applied to the TCE PBPK model outlined in Section 2. For this example, we use the two parameter estimation methods to estimate probability distributions for the fat partition coefficient $P_f$, which is expected to vary in value across a population of individuals. All other parameters were set at those values given in [1, 25]. Results for the two methods are compared using various types of simulated data that are based on different probability distributions over various lengths of time. In particular, the simulated data include concentrations in time from both the blood and fat tissue compartments that are generated using the PBPK model (2.1)–(2.13) with various levels of noise added.

In the simulation study, we generate data for concentrations of TCE in venous blood and fat, use the two methods mentioned with this data to recover the partition coefficient for fat $P_f$ and compare the results. We generate data for $n = 10$ individuals, each with $n_i = 31$ observations, equally spaced from time 0 to 2 hours and 0 to 5 hours, respectively. We add different levels and types of noise to the concentrations. The relative noise levels are 1 %, 5 % and 10 %, using a normal distribution and using a uniform distribution, respectively. Note that this means that for normally distributed noise, we use 0.0033, 0.0167 and 0.0333 as the standard deviation, while for uniformly distributed noise, we bound the value to be within $\pm 1\,\%$, $\pm 5\,\%$, $\pm 10\,\%$, respectively. In a set of examples, we seek distributions for $P_f$ with sets of data generated by a delta function, a normal and a mixture of two normals, respectively.

We estimate $P_f$ using the parametric approach and the nonparametric approach. We use the MCMC method implemented in MCSim for the parametric approach, assuming a model in which the distribution of the $P_{f_i}$'s are normally distributed with unknown mean and variance. The priors we assume for the mean and variance of the $P_{f_i}$'s are uniform from 1 to 100 and inverse-Gamma with coefficients 0.5, 0.5 respectively. The same inverse gamma priors are used for the noise variances. We use the PMF method implemented in Matlab for the nonparametric approach

Results from 78 different example calculations are given in [7]. We present several of these here and summarize our overall findings. First we present results from an example employing simulated data generated with a Dirac delta function as the "true" distribution. In Figures 2 and 3 we compare the parametric and the nonparametric results in the presence of 10 % relative normal noise added to the "measurements". These results are typical of our findings with this example, with the results being relatively unchanged whether we use "data" from $[0, 2]$ hours or from $[0, 5]$ hours. As might be expected, the parametric method produces normal estimates with the variances increasing as the noise level increases. Moreover, the posterior estimates have more variance (more than twice as much) when uniformly distributed noise is added to the data in place of normally distributed noise.

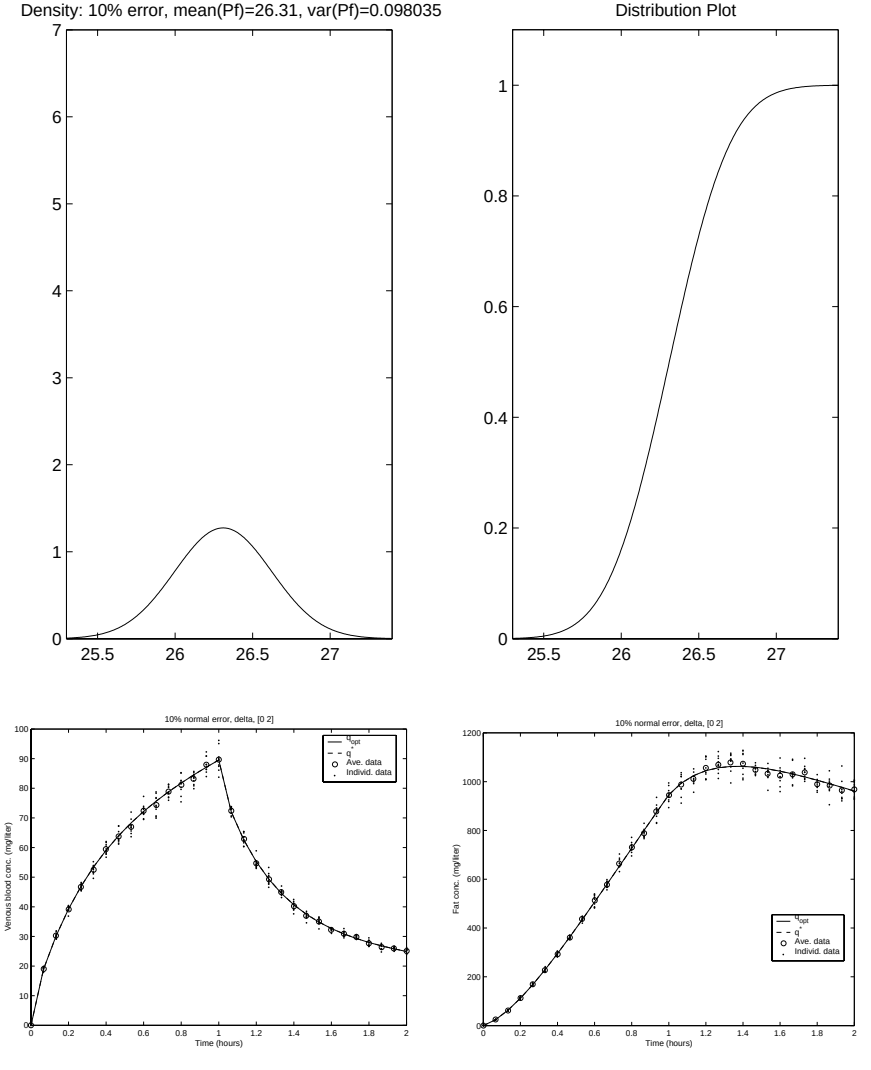


Figure 2. Parametric method. The generating density is a Dirac delta function with atom at 26.26; 10 % normal measurement noise. Top left: The estimated density function. Top right: The estimated distribution function. Bottom left: The estimated time course plots of concentration in blood. Bottom right: The estimated time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

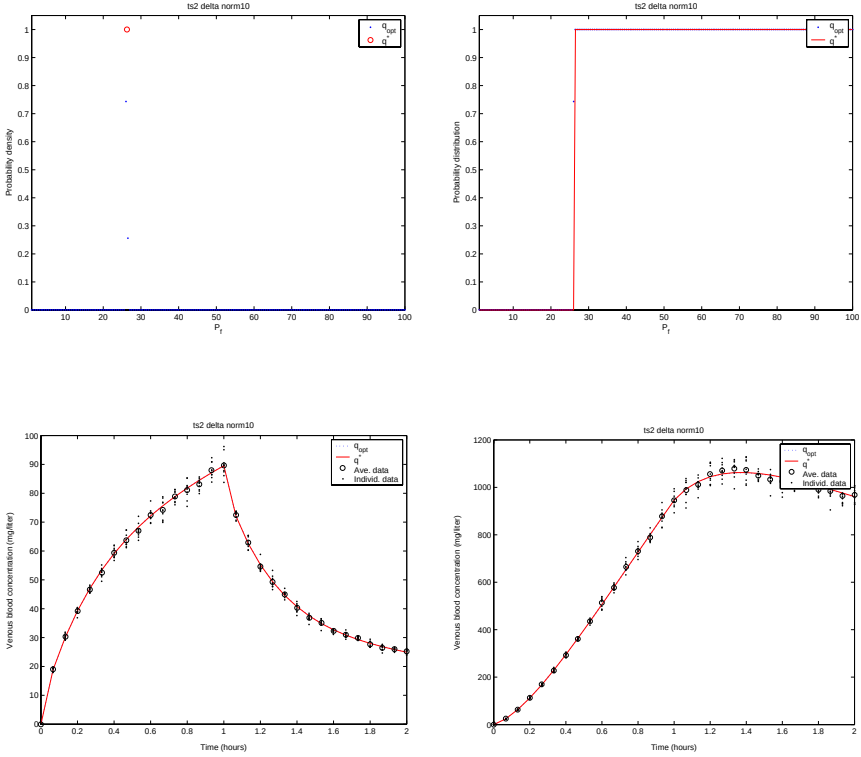


Figure 3. Nonparametric method. The generating density is a Dirac delta function with atom at 26.26; 10 % normal measurement noise. Top left: The estimated (small dots) and the data-generating (open circle) density functions. Top right: The estimated (small dots, indistinguishable except at the jump) and the data-generating (step function with one discontinuity) distribution functions. Bottom left: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in blood. Bottom right: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

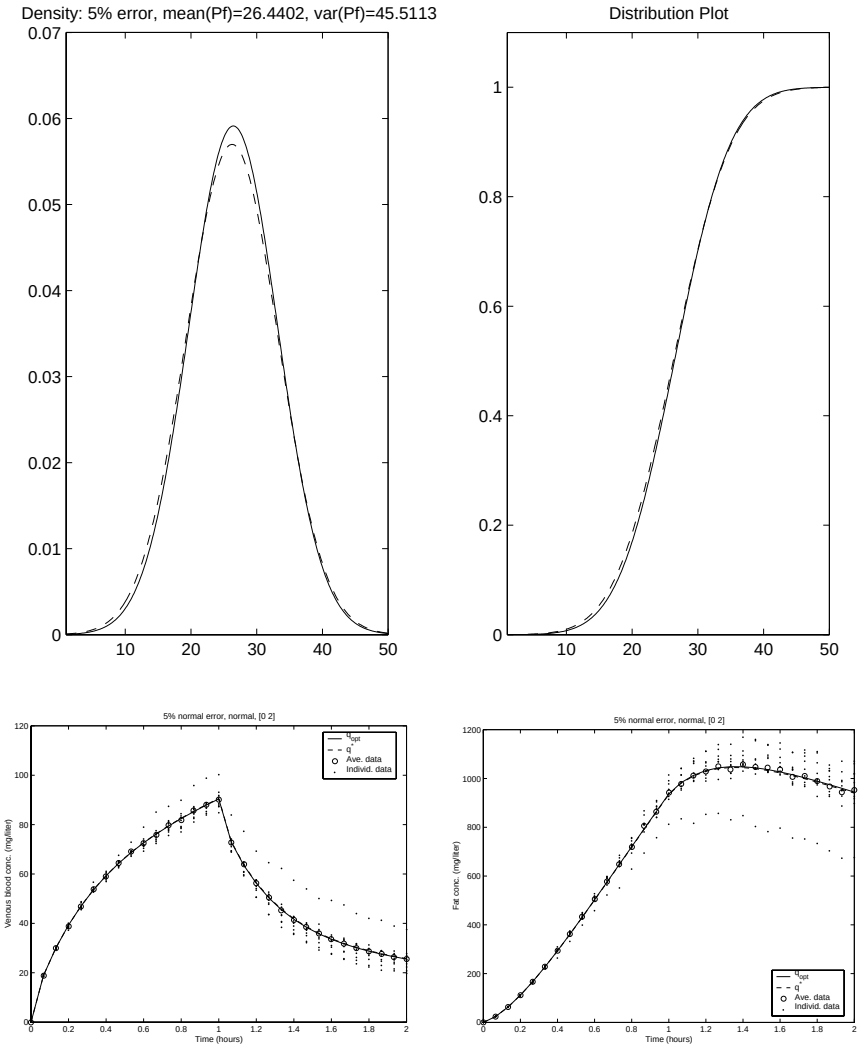


Figure 4. Parametric method. The generating density is a normal distribution with mean 26.26 and variance 49; 5% normal measurement noise. Top left: The estimated (solid line) and the data-generating (dashed line) density functions. Top right: The estimated (solid line) and the data-generating (dashed line) distribution functions. Bottom left: The estimated (solid line) and the data-generating (dashed line) time course plots of concentration in blood. Bottom right: The estimated (solid line) and the data-generating (dashed line) time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

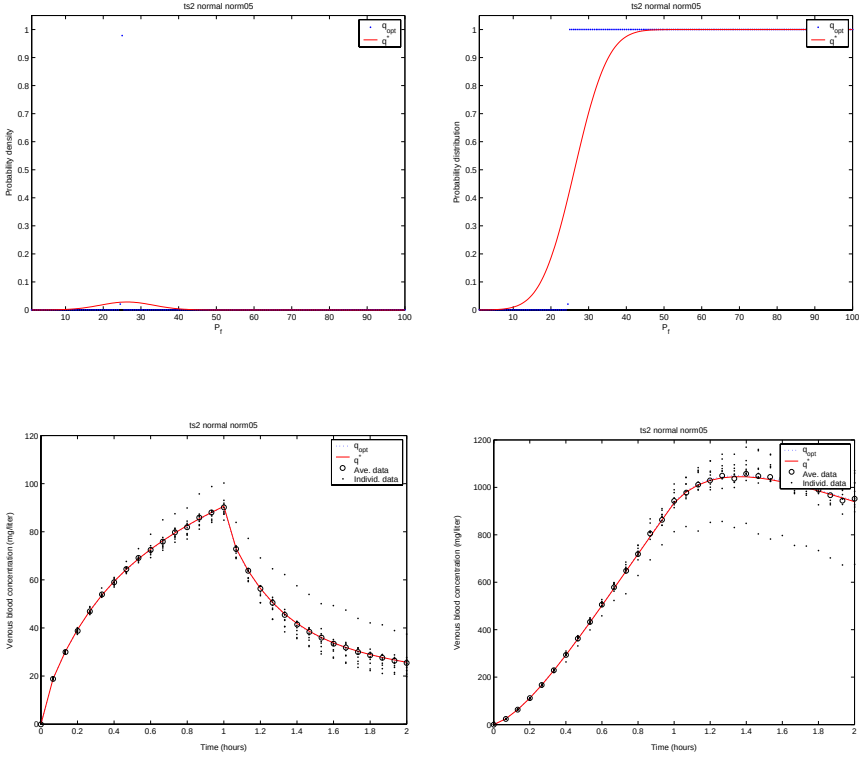


Figure 5. Nonparametric method. The generating density is a normal distribution with mean 26.26 and variance 49; 5 % normal measurement noise. Top left: The estimated (small dots) and the data-generating (solid line) density functions. Top right: The estimated (broken segments) and the data-generating (solid line) distribution functions. Bottom left: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in blood. Bottom right: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

Figures 4 and 5 depict typical findings when the data generating distribution is a unimodal normal distribution; in these examples the data have 5% relative normally distributed noise in the "measurements". Again, the results are essentially insensitive to whether the observations are taken over $[0, 2]$ vs. $[0, 5]$ hours, or whether normally or uniformly distributed noise is present in the data. Note that in the top left of Figure 5, the data generating density function is the same as in Figure 4 (with a maximum of approximately 0.055 while the estimated distribution has atoms with probabilities 0.02, 0.98). In the scale necessary to depict both graphs, the generating density is very small. Similar remarks hold for Figure 7 (atoms with probability 0.49, 0.51), Figure 8 (probabilities 0.16, 0.56 and several smaller) and Figure 9 (probabilities 0.39, 0.61). Recall also that *density* convergence is not guaranteed in the PMF approach.

In Figures 6 through 9 we present a sample of results using the two methods with data generated using a bimodal normal distribution. For the parametric method, there is little difference in the results whether one adds small (1%) or large (10%) relative noise to the data, whether it is normally or uniformly distributed noise, and whether one takes observations on $[0, 2]$ hours or $[0, 5]$ hours. The results obtained using the parametric method in all these cases do not differ substantially from those depicted in Figure 6. However, increased levels of noise in the data do degrade the ability of the nonparametric method to yield results that clearly suggest the presence of a bimodal distribution as the "true" distribution (compare Figures 7 and 8). This can be partially compensated for by taking data over a longer period (compare Figures 8 and 9 and see also the examples in [9]). However, it is rather clear that a smoother family of approximations (for example, piecewise linear or cubic splines as mentioned in Section 4 and discussed in [8]) would be more appropriate than the sum of Dirac approximations of (4.3) in examples such as these wherein one is attempting to estimate a distribution possessing a smooth density function.

In summarizing, we see that for the normal distribution examples, the parametric method provides much better results than the nonparametric method. But as soon as the true distribution departs from normality, the parametric method is either inferior or fails. In the mixture of normals example, the parametric method can recover a mean and variance but not the bimodal character of the true distribution. However, the nonparametric method is able to detect that there are two modes in the distribution.

We note with interest that although the estimated distribution is far from true in the parametric approach when an incorrect model is specified, the estimation of the mean and variance of the unknown distribution is, in fact, not affected very much by the choice of the model. This has been observed by other investigators. Whether it is true that the estimation of the first two moments of the population distribution is insensitive to the choice of the population model remains an interesting and challenging question.

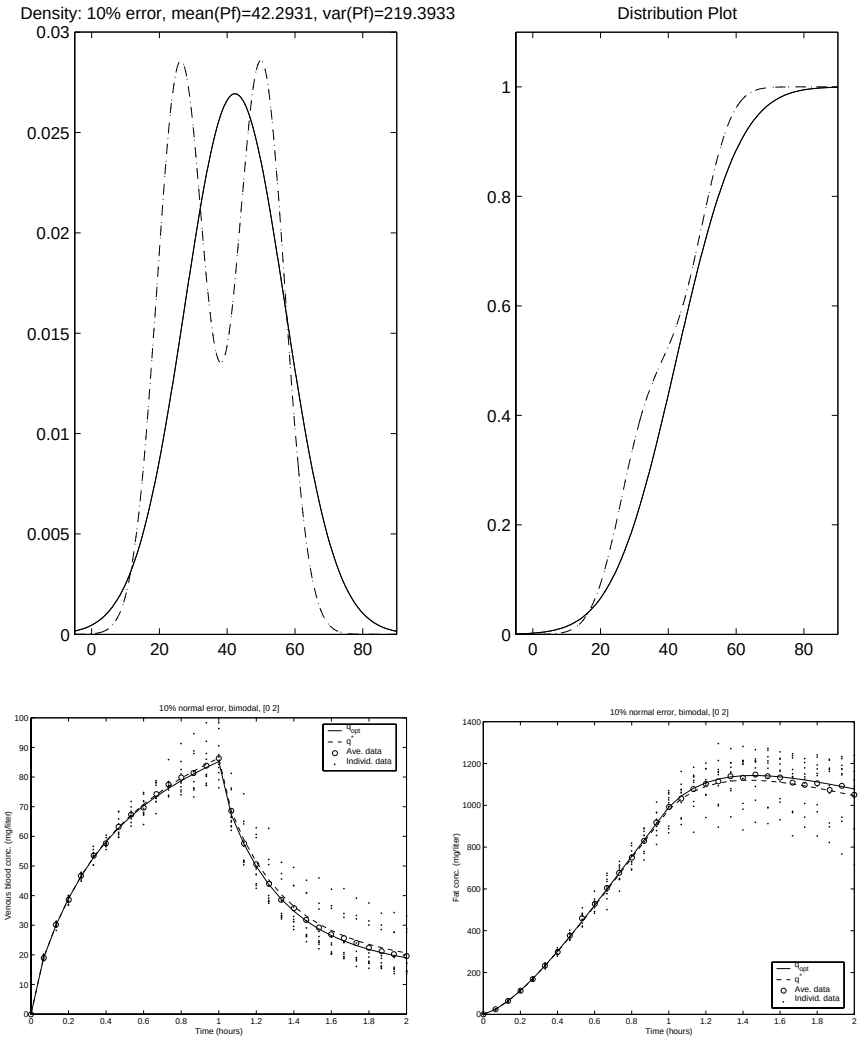


Figure 6. Parametric method. The generating density is a bimodal formed by a mixture of two normal distributions, each with mean and variance (26.26, 49) and (50,49) respectively; 10 % normal measurement noise. Top left: The estimated (solid line) and the data-generating (dashed line) density functions. Top right: The estimated (solid line) and the data-generating (dashed line) distribution functions. Bottom left: The estimated (solid line) and the data-generating (dashed line) time course plots of concentration in blood. Bottom right: The estimated (solid line) and the data-generating (dashed line) time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

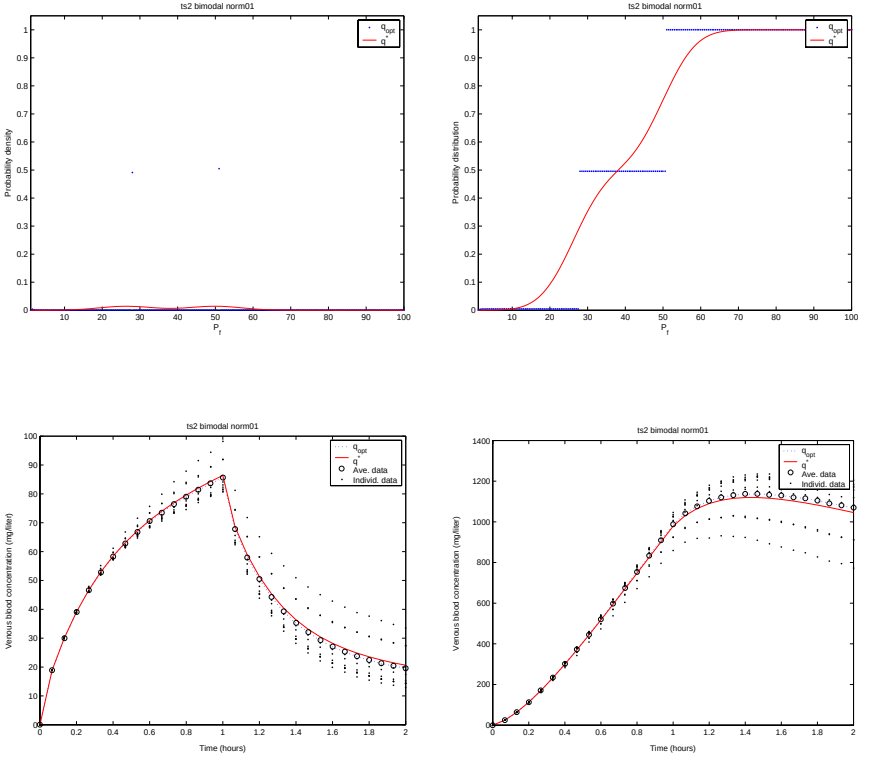


Figure 7. Nonparametric method. The generating density is a bimodal formed by a mixture of two normal distributions, each with mean and variance (26.26, 49) and (50,49) respectively; 1% normal measurement noise. Top left: The estimated (small dots) and the data-generating (solid line) density functions. Top right: The estimated (broken segments) and the data-generating (solid line) distribution functions. Bottom left: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in blood. Bottom right: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

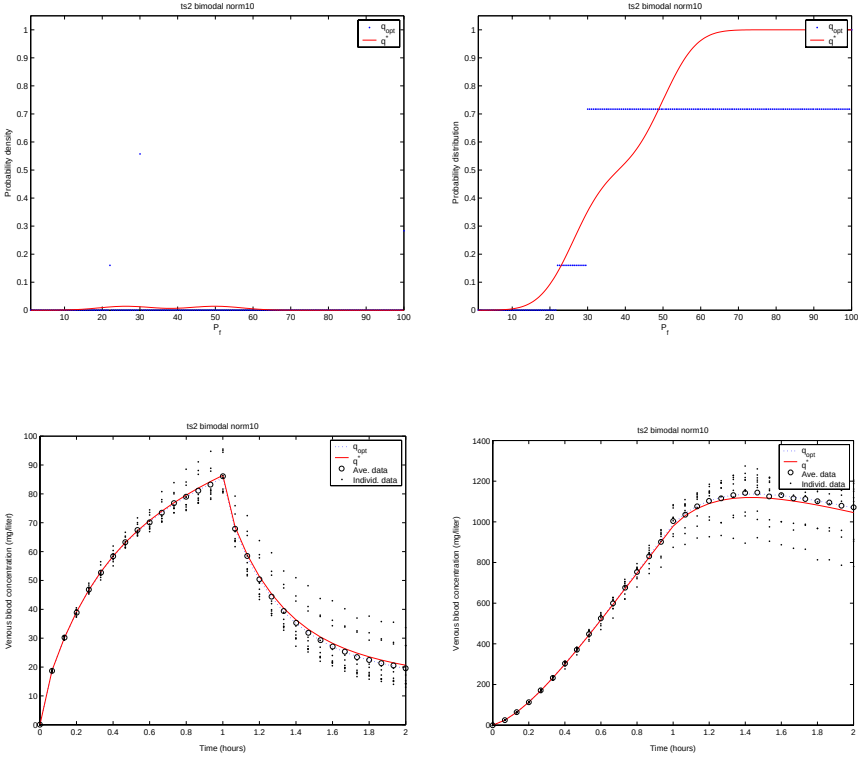


Figure 8. Nonparametric method. The generating density is a bimodal formed by a mixture of two normal distributions, each with mean and variance (26.26, 49) and (50,49) respectively; 10% normal measurement noise. Top left: The estimated (small dots) and the data-generating (solid line) density functions. Top right: The estimated (broken segments) and the data-generating (solid line) distribution functions. Bottom left: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in blood. Bottom right: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

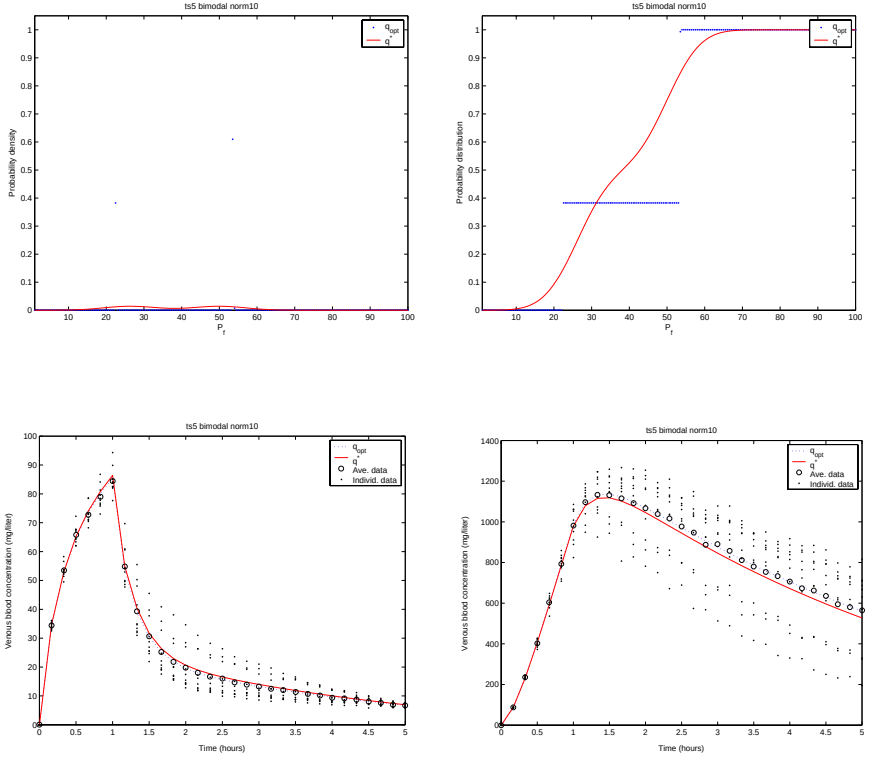


Figure 9. Nonparametric method. The generating density is a bimodal formed by a mixture of two normal distributions, each with mean and variance (26.26, 49) and (50,49) respectively; 10% normal measurement noise. Top left: The estimated (small dots) and the data-generating (solid line) density functions. Top right: The estimated (broken segments) and the data-generating (solid line) distribution functions. Bottom left: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in blood. Bottom right: The estimated (dotted line) and the data-generating (solid line) time course plots of concentration in fat. Individual data are plotted using dots, the average of the observations is represented using circles in the bottom figures

## 6. CONCLUSIONS

In this paper we have compared an MCMC parametric approach to a nonparametric method for the estimation of distributions in a hierarchical setting. Extensive computations using simulated data generated with a number of different underlying known distributions (Dirac delta, unimodal normal and bimodal mixture of normals) were carried out with a typical PBPK model for individuals. Based on the computations summarized here and those found in [9], we can make some comments about the advantages and disadvantages of the two methods.

We first note that the parametric method has the following advantages: (i) It is relatively easy to incorporate prior knowledge about uncertainty directly into an MCMC approach; (ii) Population level structures are readily incorporated onto individual level models in a hierarchical setting; (iii) If the prior distribution assumptions are accurate, the method generally provides very good fits to the data and the underlying uncertainty in the data. On the other hand, disadvantages can be substantial and include: (i) One must impose an underlying structure at the population level, which, because of the criticality to the success of the method, can provide completely misleading results if incorrectly chosen; (ii) It is not always easy to recognize when an incorrect prior assumption has led to an incorrect posterior estimate, and thus one may readily interpret convergence to the incorrect distribution as a successful estimation; (iii) Implementation of MCMC is not trivial when the mathematical model is specified implicitly through differential equations. MCSim does have an ODE solver incorporated in the software but if the system dynamics are given by a PDE model, semi-discretization to an ODE approximating system generally leads to massive memory difficulties; (iv) Finally, the computations using MCMC can be quite time consuming. For example, in comparing our examples in a typical MCMC parametric and a typical PMF nonparametric calculation (see [7]), the results with MCSim (for only 1000 samples generated) required 466 seconds while the corresponding PMF calculations took only 215 seconds. This time requirement for MCMC is a serious drawback in examples with more complex system dynamics.

Advantages for the nonparametric PMF approach include: (i) One is not required to provide an initial distribution structure at the population level; (ii) The method provides estimation capabilities even when the sought-after distribution is not similar to any known distribution; (iii) Implementation even in the context of a hierarchical setting leads to a standard constrained quadratic programming problem for which excellent software is widely available; (iv) If one does impose a prior distributional structure, the method, as described in Section 4, readily becomes a parametric method that is computationally efficient; (v) If one suspects a relatively smooth underlying distribution, higher order spline methods for approximations can be used in place of the Dirac approximations introduced in Section 4. This flexibility typically provides better estimation results in such problems. All this being said, the method can be computationally challenging and can fail to converge if certain ill-conditioning

is present in the associated approximating quadratic programming problems. Moreover, one is not guaranteed convergence of the approximating densities in any specific topology (or perhaps none at all).

The problems we are addressing are quite difficult when multiple parameters (distributions) are to be estimated. In such cases, both approaches can encounter serious computational difficulties. There are many open and interesting theoretical and computational questions yet to be investigated in this area of research.

## REFERENCES

1. R. A. Albanese, H. T. Banks, M. V. Evans, and L. K. Potter, Physiologically based pharmacokinetic models for the transport of trichloroethylene in adipose tissue. *Bulletin of Mathematical Biology* (2002) **64**, 97–131.

2. H. T. Banks, Remarks on uncertainty of assessment and management in modeling and computation. *Mathematical and Computer Modeling* (2001) **33**, 39–47.

3. H. T. Banks and K. L. Bihari, Modeling and estimating uncertainty in parameter estimation. *Inverse Problems* (2001) **17**, 95–111.

4. H. T. Banks, L. W. Botsford, F. Kappel, and C. Wang, Modeling and estimation in size structured population models. In: *Proc. 2nd course on Math. Ecology.* World Scientific Press, Singapore, 1988, 521–541.

5. H. T. Banks, B. G. Fitzpatrick, L. K. Potter, and Y. Zhang, Estimation of probability distributions for individual parameters using aggregate population data. In: *Stochastic Analysis, Control, Optimization and Applications: a Volume in Honor of W. H. Fleming.* Birkhäuser, Boston, 1999, 353–371.

6. H. T. Banks and B. G. Fitzpatrick, Estimation of growth rate distributions in size-structured population models. *Quart. Appl. Math.* (1991) **49**, 215–235.

7. H. T. Banks, Y. Ma, and L. K. Potter, A simulation-based comparison between parametric and nonparametric estimation methods in PBPK models. *CRSC-TR04-25, June, 2004.*

8. H. T. Banks and G. A. Pinter. A probabilistic multiscale approach to hysteresis in shear wave propagation in biotissue. *CRSC-TR04-03, January, 2004*; *SIAM J. Multiscale Modeling and Simulation* (submitted).

9. H. T. Banks and L. K. Potter, Probabilistic methods for addressing uncertainty and variability in biological models: Application to a toxicokinetic model. *CRSC-TR02-27, September, 2002*; *Math. Biosci.* (submitted).

10. H. T. Banks, L. K. Potter, and Y. Zhang, Use of aggregate size-structured population data to estimate distribution of growth rates. In: *Memoria 8th Int. Congress on Biomathematics, Panama '97, August.* 1997, 3–12.

11. J. Besag, Spatial interaction and the statistical analysis of lattice systems. *Journal of Royal Statistical Society, Series B* (1974) **55**, 25–37.

12. M. Davidian and D. Giltinan, *Nonlinear Models for Repeated Measurement Data.* Chapman & Hall, London, 1998.

13. A. Gelman, F. Bois, and J. Jiang, Physiological pharmacokinetic analysis using population modeling and informative prior distributions. *Journal of the American Statistical Association* (1996) **91**, 1400–1412.

14. W. Gilks, S. Richardson, and D. Spiegelhalter, *Markov Chain Monte Carlo in Practice.* Chapman & Hall/CRC, New York, 1996.

15. W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* (1970) **57**, 97–109.

16. J. Kaipio, V. Kolehmainen, E. Somersalo, and M. Vauhkonen, Statistical inversion and Monte Carlo sampling methods in electrical impedance tomography. *Inverse Problems* (2000) **16**, 1487–1522.

17. N. Lange, B. Carlin, and A. Gelfand, Hierarchical Bayes models for the progression of HIV infection using longitudinal CD4 T-cell numbers. *Journal of American Statistical Association* (1992) **87**, 615–626.

18. B. G. Lindsey, The geometry of mixture likelihoods: a general theory. *Annals of Statistics* (1983) **11**, 86–94.

19. B. G. Lindsey, Mixture models: theory, geometry and applications. In: *NSF-CBMS Regional Conf. Series in Prob. and Statistics. Vol. 5.* Inst. Math. Stat, Haywood, CA, 1995.

20. A. Mallet, A maximum likelihood estimation method for random coefficient regression models. *Biometrika* (1986) **73**, 645–656.

21. M. A. Medinsky and C. D. Klaassen, *Toxicokinetics.* Casarett and Doull's Toxicology: The Basic Science of Poisons. McGraw-Hill, Health Professions Division, New York, 5th edition, 1996.

22. N. Metropolis, A. W. Rosenbluth, M. M. Rosenbluth, A. H. Teller, and E. Tellet, Equations of state calculations by fast computing machines. *Journal of Chemical Physics* (1953) **21**, 1087–1091.

23. K. Mosegaard and M. Sambridge, Monte Carlo analysis of inverse problems. *Inverse Problems* (2002) **18**, R29–R54.

24. E. Nummelin, *General Irreducible Markov Chains and Non-Negative Operators.* Cambridge University Press, Cambridge, 1984.

25. L. K. Potter, *Physiologically Based Pharmacokinetic Models for the Systemic Transport of Trichloroethylene.* PhD thesis, North Carolina State University, Raleigh, NC, August 2001. `www.lib.ncsu.edu.`

26. A. Schumitzky. The nonparametric maximum likelihood approach to pharmacokinetic population analysis. In: *Proceedings of the Western Simulation Multiconference — Simulation in Health Care, Society for Computer Simulation*, 1993.

27. L. Tierney, Markov Chains for exploring posterior distributions. *Annals of Statistics* (1970) **22**, 1701–1762.

28. J. Wakefield, The Baysian analysis of population pharmacokinetic models. *Journal of American Statistical Association* (1996) **91**, 62–72.