

Parameter Estimation Using Unidentified Individual Data in Individual Based Models

H.T. Banks *, R. Baraldi, J. Catenacci, N. Myers

Center for Research in Scientific Computation, North Carolina State University
Raleigh, NC 27695-8212 USA

Abstract. In physiological experiments, it is common for measurements to be collected from multiple subjects. Often it is the case that a subject cannot be measured or identified at multiple time points (referred to as unidentified individual data in this work but often referred to as aggregate population data [5, Chapter 5]). Due to a lack of alternative methods, this form of data is typically treated as if it is collected from a single individual. This assumption leads to an overconfidence in model parameter values and model based predictions. We propose a novel method which accounts for inter-individual variability in experiments where only unidentified individual data is available. Both parametric and nonparametric methods for estimating the distribution of parameters which vary among individuals are developed. These methods are illustrated using both simulated data, and data taken from a physiological experiment. Taking the approach outlined in this paper results in more accurate quantification of the uncertainty attributed to inter-individual variability.

Keywords and phrases: pharmacokinetics, pharmacodynamics, physiological based pharmacokinetic modeling, inter-individual parameter variability, uncertainty quantification, random differential equations, distribution estimation, Prohorov metric framework, aggregate population data.

Mathematics Subject Classification: 92C45, 34F05, 49N45

1. Introduction

Pharmacokinetics (PK) is the study of how an administered drug is absorbed, distributed, metabolized, and excreted over time. Pharmacodynamics (PD) quantifies the relationship between drug exposure and the resulting biochemical and physiological effects. Physiologically based pharmacokinetic (PBPK) models have been developed for use in areas such as estimating internal doses, extrapolating across populations, quantifying pharmacokinetic parameters, understanding drug specific metabolic pathways, and designing experiments. In physiological experiments, often one does not study an individual but rather a collective group of individuals, which we will refer to as a population. Individuals constitute the subject of interest in the experiments which are described by an associated mathematical model. For example, in cell growth models where the data are cell counts, the individual is the independently grown cell colonies. In ecological

*Corresponding author. E-mail: htbanks@ncsu.edu

studies where the subject's attributes are measured (i.e., size, weight, age, etc.), the individual refers to the individual subject or specimen that is measured. In pharmacology studies where the drug concentration is of interest, the individuals are the subjects to which the drug was administered (see the mouse organ analyses example in Section 4 below). Further complicating the scenario is that for many applications it is not always possible to track the *same* individuals over time. This is precisely the case in pharmacological experiments where the subjects must be sacrificed in order to measure the quantity of interest. When only data of this form is available, the current practice is to treat the data as if it was collected from a single individual over time (i.e., individual data).

Mathematical models are typically developed in PK/PD efforts to describe the dynamics of a single individual as opposed to describing the population. Alternative approaches can be ubiquitously found in other fields such as cell biology, ecology, and in many physics based applications (at the macroscopic level) where the population behavior is modeled directly rather than describing each individual, cell, organism, or particle. The mathematical models in PK/PD applications are of course dependent on a set of parameters. We will divide the types of parameters into two classes: the parameters that are constant across all members in the population we will refer to as the *population level parameters*, and the parameters that vary significantly from one individual to the next we will refer to as *individual parameters*. The practice of using sets of longitudinal data from a number of different individuals to estimate the model parameters (both individual and population level parameters) has been well studied. This methodology is typically referred to as population or multilevel models, and often one wishes to ascertain information about the underlying distribution which best describes the individual parameters. To the authors' knowledge, the first use of a multilevel model in a pharmacokinetic study was made by Sheiner in 1984 [22]. A recent review article [7] provides a good summary of several common techniques for the case of multilevel models, and specific work on non-linear mixed effects models can be found in [11, 17, 20]. Additionally, Bayesian approaches to model calibration and data assimilation are becoming more popular in the area of PK/PD modeling. A nice example of using a Bayesian approach to estimate the inter-individual variability among relevant model parameters is given in [15], and a recent survey article [12] provides an overview of various stochastic methods for the modeling and estimation in PK/PD models. However, each of these approaches requires that one has longitudinal data from the same individual subjects. That is, measurements are taken from the same individual over time. Here, we are interested in the case where we have longitudinal data obtained from individual subjects in which the subjects cannot be identified or the data is not necessarily obtained from the same subjects over time. Experimentally this is a result of either not being able to test the same subject repeatedly (common in many cell growth, chemical, and pharmacology experiments), or if the subject from which the data was collected cannot be identified (typically the scenario arising in ecological modeling, and often referred to as aggregate data in the ecological community).

In this work, we will investigate how to use longitudinal data collected from unidentified individuals to estimate the population level parameters as well as the *distribution* of the individual parameters. The importance of capturing the parameter variability among individuals has long been recognized, and recently it has been shown that different genotypes can play a significant role on how a drug is metabolized and processed [13, 18]. Thus, capturing the inter-individual variation in model parameters is crucial, especially when considering extrapolation between species, doses, and exposure routines. As mentioned previously, several methods have been developed specifically for the case of estimating the distribution of individual parameters so long as longitudinal individual data is available. Yet there are a lack of methods available aimed at accounting for parameter variability when true individual data is not available. To the authors' knowledge, there is only a little previous work (see [1, 4, 8]) in the mathematical or pharmacological literature which attempt to account for inter-individual variability when individual subjects cannot be tracked over time. When aggregate data is treated as individual data, as is traditionally done in most physiological pharmacokinetic studies, then one can only account for the average dynamics, and all parameter values (either estimated through means of an inverse problem, or measured in separate experiments) represent the *average* behavior. There has been a recent push in the pharmacology modeling

community to better understand the uncertainty and variability associated with the model development and calibration, and incorporating this information into the risk assessment process (see [9, 10, 19] and the references therein). We emphasize that capturing the uncertainty due to inter-individual variability when the individual is not available for repeated measurements has yet to be addressed. We propose that the methods outlined in this work provide a more realistic quantification of parameter uncertainty in the case where only unidentified individual data is available, which in turn leads to more reliable predictive capabilities and risk assessment analysis.

The remainder of the paper is outlined as follows. In Section 2 we begin by providing the mathematical framework necessary to address inter-individual variability using both parametric and nonparametric approaches. In Section 3 an example using the logistic equation is presented which demonstrates the error encountered when failing to account for inter-individual variability. We also demonstrate the steps necessary to implement both the parametric and nonparametric version of our method. The application of our methods to an existing experimental PBPK data set is investigated in Section 4, and we conclude by summarizing our findings in Section 5.

2. Model description

Assume that data are available at time points t_j from $i = 1, \dots, I_j$ individuals, $j = 1, \dots, n_t$. It is paramount to clarify at this point that we are assuming that we *cannot* distinguish individuals over time. Furthermore, we note that we are not assuming that the same number of individuals are measured at each time (thus the additional subscript I_j). That is, at time t_j , data is collected from I_j subjects, but these subjects cannot be identified as those from whom data was collected at any other time. Thus, at time t_{j+1} , data is collected from I_{j+1} subjects and these subjects cannot be identified to correspond to any of the subjects that provided data at time t_j .

We denote the mathematical model describing subject i by

$$\begin{aligned} \frac{dx^{(i)}}{dt} &= f(t, x^{(i)}; \theta, \phi^{(i)}), & x^{(i)}(0) &= x_0, \\ \phi^{(i)} &\sim_{i.i.d.} p_\phi \end{aligned} \tag{2.1}$$

where $\theta \in \mathbb{R}^p$ are the population parameters, $\phi^{(i)} \in \mathbb{R}^s$ are the individual parameters, $x^{(i)} \in \mathbb{R}^d$, and $f : [0, \infty] \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is Lipschitz continuous. Observe that we have assumed that the individual parameters are independent and identically distributed (*i.i.d.*) random variables with density p_ϕ . Hence, equation (2.1) is a random differential equation [5, 23, 24].

We assume that the following statistical model can describe the observations

$$y_j^{(i)} = g^{(i)}(t_j; \theta_0, \phi_0^{(i)}) + \epsilon_{ij}, \quad i = 1, \dots, I_j, \quad j = 1, \dots, n_t, \tag{2.2}$$

where g is the observed part of the solution to the mathematical model, θ_0 and $\phi_0^{(i)}$ are the “true” or nominal parameter values, and ϵ_{ij} are *i.i.d.* measurement errors with mean 0 and constant variance σ_ϵ^2 .

Our goal in this work is to formulate a mathematical framework in which the observations $y_j^{(i)}$ are utilized to estimate the population level parameters θ as well as describe the variability of the individual parameters $\phi^{(i)}$. We will quantify the variability in the individual parameters by either computing the first two moments of the density p_ϕ , or estimating the density directly. For a given application, if estimating the first two moments of p_ϕ provides sufficient information about the variability in the individual parameters, then one can take a parametric approach. However, if estimation of the full density p_ϕ is required, then one should choose a nonparametric approach if possible (unless there is a high level of certainty about the form of p_ϕ).

We remark that one can obtain a partial differential equation describing the time evolution of the probability density function of X denoted by $p_x(t, x)$. First let $z = [x^T \ \phi^T]^T$. Then consider the system

$$\frac{dz}{dt} = \tilde{f}(t, z), \quad z(0) = [X_0^T \ \Phi^T]^T, \tag{2.3}$$

where $\tilde{f}(t, z) = [f(t, x)^T \ 0]^T$. Then, by Liouville's equation (see [5, Chap. 7.3], the joint probability density function $u_{X, \Phi}(t, x, \phi)$ of X and Φ satisfies

$$\frac{\partial}{\partial t} u_{X, \Phi}(t, x, \phi) + \sum_{k=1}^d \frac{\partial}{\partial x_k} (f_k(t, x) u_{X, \Phi}(t, x, \phi)) = 0, \quad (2.4)$$

with the initial condition $u_{X, \Phi}(0, x, \phi) = u_0(x, \phi)$, where u_0 is the joint probability density function of X_0 and Φ . Note that equation (2.4) is a first order linear partial differential equation (PDE). Given the joint probability density function of X_0 and Φ , we can obtain the probability density function for $x(t)$ given by

$$p_X(t, x) = \int_{\Omega} u_{X, \Phi}(t, x, \phi) d\phi, \quad (2.5)$$

where Ω denotes all of the possible values for ϕ .

Equation (2.4) can be rewritten in the form

$$\frac{\partial}{\partial t} u_{X, \Phi}(t, x, \phi) + \sum_{k=1}^d f_k(t, x) \frac{\partial}{\partial x_k} u_{X, \Phi}(t, x, \phi) = c(t, x, \phi) u_{X, \Phi}(t, x, \phi), \quad (2.6)$$

where

$$c(t, x, \phi) = - \sum_{k=1}^d \frac{\partial}{\partial x_k} f_k(t, x).$$

The characteristic equations of (2.6) are given by

$$\frac{dx_k}{dt} = f_k(t, x), \quad \frac{d\phi_l}{dt} = 0, \quad \frac{du}{dt} = c(t, x, \phi)u, \quad (2.7)$$

with the initial conditions

$$x_k(0, \xi, \rho) = \xi_k, \quad \phi_l(0, \xi, \rho) = r_l, \quad u(0, \xi, \rho) = u_0(\xi, \rho), \quad (2.8)$$

where $k = 1, \dots, d$, and $l = 1, \dots, s$.

Definition 2.1. Consider equation (2.4) with the initial conditions (2.8) defining an initial curve. The equation and the initial curve are said to satisfy the *transversality condition* at a point (ξ, ρ) on the initial curve, if $J|_{t=0} \neq 0$, where $J = \left| \frac{\partial(x, \phi)}{\partial(\xi, \rho)} \right|$ is the Jacobian matrix.

Theorem 2.2. Assume that $f : [0, \infty) \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ is differentiable and that $\partial f_k / \partial x_k$ are Lipschitz continuous. Assume further that the transversality condition holds at each point (ξ, ρ) in the interval $(\xi_0 - 2\delta_\xi, \xi_0 + 2\delta_\xi) \times (\rho_0 - 2\delta_\rho, \rho_0 + 2\delta_\rho) \in \mathbb{R}^{d+s}$ on the initial curve. Then (2.4) has a unique solution in the neighborhood $(t, \xi, \rho) \in (-\varepsilon, \varepsilon) \times (\xi_0 - 2\delta_\xi, \xi_0 + 2\delta_\xi) \times (\rho_0 - 2\delta_\rho, \rho_0 + 2\delta_\rho)$ of the initial curve.

It is known that a first order linear partial differential equation can be shown to have a unique solution locally, so long as the coefficients are sufficiently smooth. With the assumption that f is differentiable and that the terms $\partial f_k / \partial x_k$ are Lipschitz continuous, Theorem 2.2 follows as an immediate consequence of standard linear PDE theory.

If the model (2.1) is linear and autonomous, so that $dx/dt = A(\phi)x$, then we can easily find the solution $u(t, x, \phi)$. In this case we observe that $c(t, x, \phi) = -\text{trace}(A(\phi))$. The characteristic equations are given by

$$\frac{dx}{dt} = A(\phi)x, \quad \frac{d\phi}{dt} = 0, \quad \frac{du}{dt} = c(t, x, \phi)u, \quad (2.9)$$

with the initial conditions (2.8). We can solve (2.9) to obtain

$$x(t, \xi, \rho) = \exp(A(\phi)t)\xi, \quad \phi = \rho, \quad u(t, \xi, \rho) = u_0(\xi, \rho) \exp(-\text{trace}(A(\rho))t).$$

By inverting $x(t, \xi, \rho)$, we obtain the solution to Liouville's equation

$$u(t, x, \phi) = u_0(\exp(-A(\phi)t)x, \phi) \exp(-\text{trace}(A(\phi))t). \quad (2.10)$$

Many PK/PD models which only involve first order kinetics are linear, and thus by (2.10) combined with equation (2.5), one can readily determine the desired probability density function.

2.1. Parametric approach

In order to compute the forward simulation of the mathematical model (either through the form given in (2.1) or (2.4)) we must have a choice for p_ϕ , which of course is exactly what we are trying to recover. In this section, we will describe how to estimate the mean and variance of the distribution p_ϕ through a parametric method. It may be the case that the mean and variance will provide a sufficient description of the variability among individuals. For example, if there is evidence that p_ϕ is symmetric and unimodal, then the first two moments capture the essence of the distribution. Furthermore, as we will see in the next section, a relatively rich data set is required in estimating the density of p_ϕ . In cases where the data is sparse, one may be constrained to only estimating the first two moments. In the parametric case, we assume a form for p_ϕ , which depends on a finite set of parameters Ψ . For example, if we assume that p_ϕ is a normal distribution, then $\Psi = (\mu, \sigma^2)$, where μ and σ^2 are the mean and variance, respectively. On the other hand, if we choose a binormal distribution for p_ϕ , then $\Psi = (\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$.

Once the form of p_ϕ has been chosen, the expected value and variance (as well as higher moments) of the random differential equation (2.1) can be computed in several ways. Non-intrusive methods such as Monte Carlo sampling or stochastic collocation (SC) provide fairly straightforward implementation. Another popular choice is a stochastic Galerkin (SG) approach using the generalized polynomial chaos (gPC) framework. Recently, there have been a plethora of articles published on various SC and SG methods. Choosing a suitable and efficient method for simulating a set of random differential equations is problem dependent and beyond the scope of this work. We direct the interested reader to [25], which provides a thorough explanation of the fundamental basis for gPC methods.

For simplicity, we assume $g^{(i)}(t_j; \theta, \phi^{(i)}) = x^{(i)}(t_j; \theta, \phi^{(i)})$. Then, given a set of data $y_j^{(i)}$, estimates for the population and individual parameters can be obtained by solving

$$(\hat{\theta}, \hat{\Psi}, \hat{\sigma}_\epsilon) = \underset{(\theta, \Psi, \sigma_\epsilon) \in \Theta}{\text{argmin}} \mathcal{J}(\theta, \Psi, \sigma_\epsilon), \quad (2.11)$$

where the cost functional is given by

$$\begin{aligned} \mathcal{J}(\theta, \Psi, \sigma_\epsilon) = & \sum_{j=1}^{n_t} \left(\mathbb{E}[y_j^{(i)}] - \mathbb{E}[x(t_j; \theta, \Psi)] \right)^2 \\ & + \left(\sqrt{\text{Var}[y_j^{(i)}]} - \sqrt{\text{Var}[x(t_j; \theta, \Psi)] + \sigma_\epsilon^2} \right)^2, \end{aligned} \quad (2.12)$$

and Θ is the space of admissible parameters which is assumed to be compact. The results of this inverse problem provide simultaneously estimates for the population parameter θ as well as the parameters Ψ for the distribution $p_\phi(\Psi)$ from which the moments can readily be determined. Notice that we have included in the inverse problem the estimation the variance of the measurement errors σ_ϵ^2 . This is required in order to obtain an unbiased estimate for the variance parameters of Ψ (assuming that ϵ_j and x_j are independent).

Even if we take a standard maximum-likelihood estimation (MLE) approach (or the equivalent Bayesian approach) in which we treat the data as individual data (i.e., as if the data was collected from a *single*

individual), we will only be able recover meaningful estimates of the mean of the individual parameters. We illustrate this point in Section 3.1. Of course, rather than using a least squares estimation procedure as is outlined above, a MLE or Bayesian approach can readily be applied to this parameter estimation problem by formulating an appropriate likelihood function analogous to (2.11)–(2.12).

2.2. Estimating the density of the individual parameters

In certain cases, it may not be enough to determine the mean and variance of the individual parameters. This is especially true if the underlying distribution is not symmetric, or is bi-modal. For these considerations, it becomes more challenging to carryout a meaningful parametric estimation. Thus, we turn to developing a nonparametric approach. However, for the nonparametric estimation, we will need more information from the data than just the time varying mean and variance. Here we require an actual density to be measured at each sampling time. We remark that the process in which we previously assumed for the data collection can be used to give approximate densities by means of kernel density estimation. The accuracy of the kernel density estimation depends upon whether or not the data points at each time t_j provide a sufficient description of the true but unknown density. This indicates that we need enough data points at every sampling time to provide a representative sampling of the density, and of course there is no way to guarantee this in practice.

In order to model the evolution of the density $p_x(x, t)$ we will make use of Liouville’s equation (2.4). In solving Liouville’s equation we must be given a probability density for the random variable Φ , as well as probability density for X_0 if the initial condition is taken as random as well. We could of course again take a parametric approach by assuming a specific form of the unknown probability density (or densities). Rather, in this case we will illustrate the more difficult task of a nonparametric estimation of the density p_ϕ . To do this we will make use of the Prohorov metric framework [2, 3, 5]. In this setting, we can build an approximation using piecewise linear splines given by

$$p_\phi^M(\phi) = \sum_{m=1}^M w_m \ell_m(\phi), \quad (2.13)$$

where $\ell_m(\phi)$ are the usual spline functions [6, 21], and where we impose the constraint

$$\sum_{m=1}^M w_m \int_{\Omega} \ell_m(\phi) d\phi = 1.$$

Given a set of weights $w = \{w_m\}_{m=1}^M$, the approximate probability density p_ϕ^M can be computed, and this in turn is used as the initial condition for (2.4). Once equation (2.4) is solved to obtain the solution $u^M(t, x, \phi)$, the approximate probability density $p_x^M(t, x)$ for $x(t)$ is given by (2.5).

The inverse problem is then given by

$$(\hat{\theta}, \hat{w}) = \underset{(\theta, w) \in \Theta}{\operatorname{argmin}} \mathcal{J}(\theta, w), \quad (2.14)$$

where the cost functional is now defined as

$$\mathcal{J}(\theta, w) = \sum_{k=1}^{n_x} \sum_{j=1}^{n_t} \left(p_{jk}(y_j^{(i)}) - p_x^M(t_j, x) \right)^2. \quad (2.15)$$

In the above equation, $p_{jk}(y_j^{(i)})$ denotes the density obtained from the data via a kernel density estimator at sampling time t_j and “spatial” discretization x_k for $k = 1, \dots, n_x$.

3. A motivating example: logistic growth

Common to many PK/PD models is the notion of a saturation limit for a quantity of interest. For example, the saturation of drug/chemical concentration, the maximum size of tumor growth, or the saturation of a population of cells. Frequently this limiting behavior results in a S-shaped curve, which can be described by a sigmoid, Hill, Gompertz, logistic, or one of several other functions.

We begin by illustrating the importance of accounting for intrinsic parameter variability with a simple logistic growth model where we assume that the growth rate varies across the individuals, but the carrying capacity (saturation limit) and initial condition are both constant over the population. This leads to the mathematical model

$$\begin{aligned} \frac{dx^{(i)}}{dt} &= \phi^{(i)} x^{(i)} \left(1 - \frac{x^{(i)}}{K} \right), & x^{(i)}(0) &= x_0 \\ \phi^{(i)} &\sim p_\phi, \end{aligned} \tag{3.1}$$

where $\phi^{(i)}$ are the (assumed *i.i.d.*) individual growth rates with density p_ϕ and $\theta = (x_0, K)^T$ are the population level parameters. Equation (3.1) is a realization of a random differential equation given by

$$\frac{dx}{dt} = \Phi x \left(1 - \frac{x}{K} \right), \quad x(0) = x_0, \tag{3.2}$$

where $\Phi \sim p_\Phi$ is a random variable.

We will assume that the growth rates are normally distributed with mean μ_Φ and variance σ_Φ^2 (that is, $\Psi = (\mu_\Phi, \sigma_\Phi)$). We simulate the data points $y_j^{(i)}$ in the following way: for $i = 1, \dots, I$, we first draw a realization for $\phi^{(i)}$ from a given distribution, then equation (3.1) is solved at the measurement time points t_j , for $j = 1, \dots, n_t$. The measurement time points were taken as $t_j = (j - 1) \frac{t_f}{n_t - 1}$ with $t_f = 5$ and $j = 1, 2, \dots, n_t$, with $n_t = 101$. Finally, we add noise drawn from a normal distribution with mean 0 and with variance σ_ϵ^2 to the simulated model solution (that is, $y_j^{(i)} = x^{(i)}(t_j) + \epsilon_{ij}$ with $x^{(i)}(t)$ the solution to the logistic model (3.1)). We thus simulate both the intrinsic variability in the growth rate and measurement noise. In Figure 1, we show an example data set with $x_0 = 10, K = 100, I = 10$ and a noise level of $\sigma_\epsilon^2 = 1$ where $\phi^{(i)}$ was drawn from a normal distribution with mean 2 and variance 0.2.

3.1. Improperly treating the data as individual data

We begin by exploring potential pitfalls when disregarding the fact that the data is indeed aggregate data, which again we emphasize is the common assumption made nearly universally in the literature. Consider a standard experimental set up where the data $y = \{y_j\}_{j=1}^{n_t}$ is collected by only collecting a single sample from a new subject at each sampling time. That is, the data is generated exactly as described in the section above with $I = 1$. Since in this section we are disregarding the fact that the measurements are collected from different individuals and instead assuming that the observations are collected from a single individual over time, we arrive at the standard statistical model given by

$$y_j = x(t_j; \theta) + \epsilon_j.$$

In the above equation, $x(t)$ denotes the solution to the (deterministic) logistic equation

$$\frac{dx}{dt} = rx \left(1 - \frac{x}{K} \right), \quad x(0) = x_0, \tag{3.3}$$

and ϵ_j is the measurement error. Clearly there is a difference between how the simulated experiment is conducted (resulting in unidentified individual data) and how the the model assumes the data is collected, that is assuming that the data is obtained from a single individual over time. We remark that this difference is in fact an example of model discrepancy (also referred to as model misspecification in the

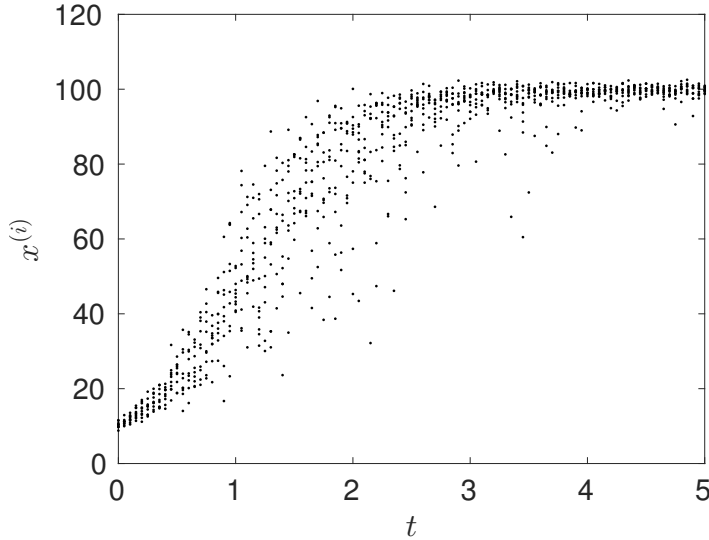


FIGURE 1. Example logistic data set with $x_0 = 10$, $K = 100$, $I = 10$ and $\sigma_\epsilon^2 = 1$. Each data point $y_j^{(i)}$ represents a sample from an unidentified individual at a given time.

literature), and the discrepancy is due to the fundamental difference between the modeling assumptions of how the data is collected and how the data is actually collected experimentally. Thus, one possible approach, which we do not explore in this work, is to attempt to account for this model discrepancy in order to arrive at calibrated parameters whose associated uncertainty agrees with the distributions of the individual parameters p_ϕ .

Bayesian estimation is a powerful tool for uncertainty analysis, and here we will illustrate that even a Bayesian procedure leads to false conclusions if one assumes that the data is collected from a single individual when in reality the data is unidentified individual data. Through the means of a Bayesian estimation, we will estimate the unknown parameters $\theta = (r, x_0, K)^T$ in the logistic model (3.3).

The prior density is denoted by $\pi_0(\theta)$, which we take as an uniform prior, and we further assume that the measurement errors are independent and identically distributed with a normal distribution having mean 0 and variance σ_ϵ^2 . With this assumption the likelihood function is given by

$$\pi(y|\theta) = \frac{1}{(2\pi\sigma_\epsilon^2)^{n_t/2}} \exp \left\{ -\frac{1}{2\sigma_\epsilon^2} \sum_{j=1}^{n_t} (y_j - x(t_j; \theta))^2 \right\}, \quad (3.4)$$

where $x(t_j; \theta)$ is the solution to the logistic equation. Then the posterior density can be obtained through

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi_0(\theta)}{\pi(y)} = \frac{\pi(y|\theta)\pi_0(\theta)}{\int_{\mathbb{R}^p} \pi(y|\theta)\pi_0(\theta)d\theta}. \quad (3.5)$$

The posterior density was approximated using the delayed rejection adaptive Metropolis (DRAM) algorithm [23]. In Figure 2, we present the model fit to a data set which was generated with the measurement error variance $\sigma_\epsilon^2 = 1$. The mean parameters were found to be

$$r = 1.9562, \quad x_0 = 11.0010, \quad K = 97.8998,$$

and we see that we are able to estimate the mean value of the growth rate, the initial condition, and the carrying capacity with reasonable accuracy. However, the approximate posterior density for the growth

rate r does not represent the true density, as can be seen in Figure 2. In fact, we severely underestimate the variance of the growth parameter r , which was estimated to be $\text{Var}[\pi(r|y)] \approx 0.0247$. This is because the posterior density represents the certainty that the given values of r describe the data set. In this framework, there is no way to distinguish between the variance due to the intrinsic variability of the growth rate and the measurement noise. The result is an (unjustifiably) high level of confidence in the parameter estimate of r , which in this case represents the mean of the true distribution, with no indication that we are not capturing the inherent variability in the system. We also show the 95% prediction interval, represented by the shaded region about the model fit in Figure 2. Clearly the prediction interval grossly underestimates the variance observed in the simulated data measurements. This is a further indication of how the overconfidence we have obtained in our parameter estimates can adversely affect the predictive capabilities in a model analysis.

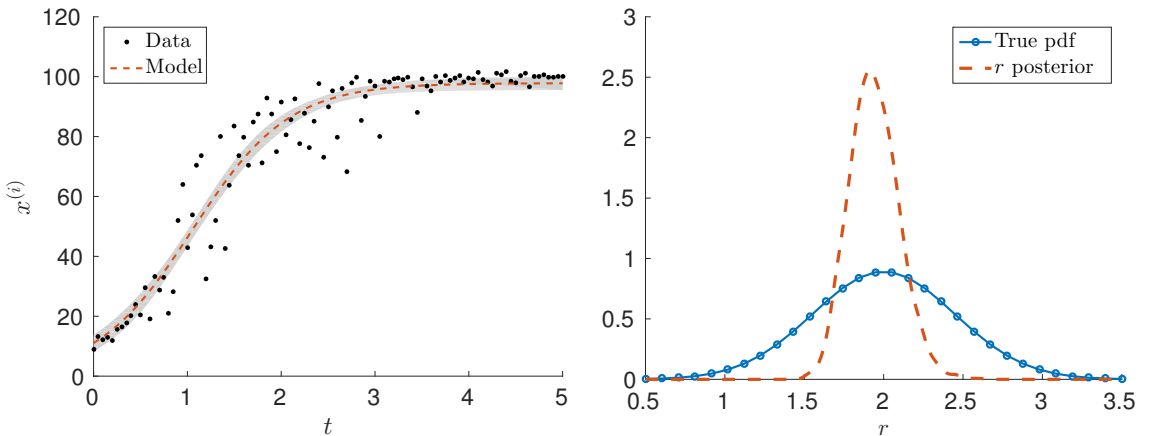


FIGURE 2. The model fit to a data set with single data point at each sampling time (left) and the posterior density of the growth rate r compared to the true density (right). The grey shaded region around the model fit is the 95% prediction interval resulting from the Bayesian estimation procedure.

3.2. Parametric estimation - model with the correct distribution

We re-emphasize that our goal is to estimate the unknown values of x_0 , K , μ_ϕ , and σ_ϕ given the data set $y_j^{(i)}$. In this first example, we take the simplest approach by correctly assuming the distribution of the individual growth rates $\phi^{(i)}$. That is, we generate our data by sampling the growth rate from a normal distribution with mean $\mu_\phi = 2$ and variance $\sigma_\phi^2 = 0.2$, and we assume in our mathematical model that the growth rate is normally distributed. The inverse problem was solved according to equations (2.11) and (2.12) where $\Psi = (\mu, \sigma)$. The moments of the model $x(t; x_0, K, \mu, \sigma)$ were obtained through a SC method. In applications, there may be either experimental evidence or biologically relevant assumptions that may guide the choice of a distribution for the individual parameters $\phi^{(i)}$. We will discuss the case of a nonparametric estimation of the unknown density p_ϕ in a later section.

A modest measurement noise of $\sigma_\epsilon^2 = 1$ is initially considered. We see from Table 1 that we are able to recover all of the parameters quite well, even with a relatively small number of individual data points at each time point. In general, the estimates improve as the number of individuals sampled at each time increases. In Figure 3, we compare the expected value of the data over time with the resulting model fit (the expected value of the RDE (3.2)), and for comparison we show the true expected value of the RDE with $\mu_\phi = 2$ and $\sigma_\phi^2 = 0.2$. The same comparison for the variance is also shown in Figure 3, where

our parameter selection allows the model to accurately account for both inter-individual variability and white noise. This results in the 95% prediction band capturing the data as desired.

I	μ_Φ	σ_Φ^2	x_0	K	σ_ϵ^2
3	2.0675	0.1660	9.6264	99.4070	0.0495
10	2.0620	0.2015	9.4836	99.8482	0.0402
30	2.0443	0.2080	9.7190	99.9118	0.1867
50	2.0476	0.2082	9.7271	99.8646	0.1706
100	2.0104	0.1970	9.9593	99.9693	0.2234
True value	2.0	0.2	10	100	1.0

TABLE 1. Estimated values for the logistic model example with measurement noise level $\sigma_\epsilon^2 = 1$. Both the simulated data and the model are assumed to have a normally distributed growth rate.

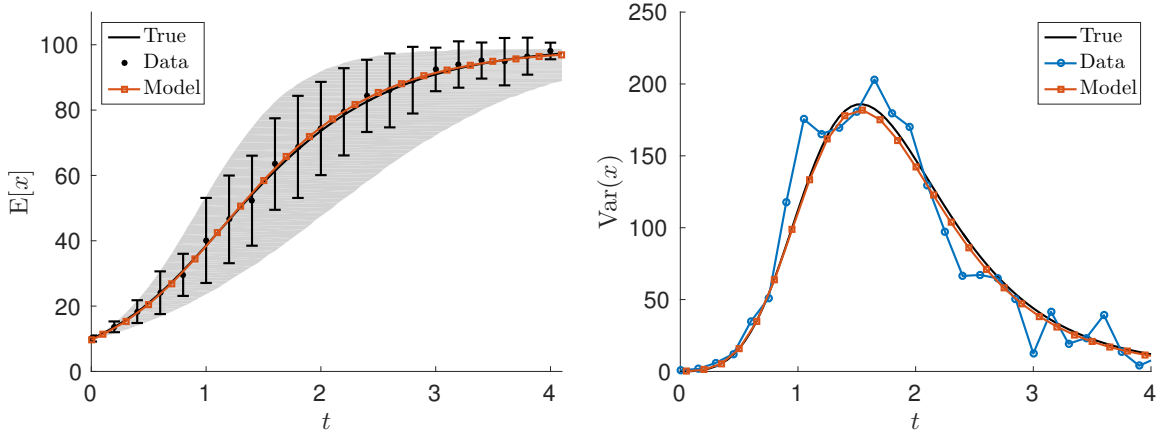


FIGURE 3. The fit of the expected value of the RDE model to the mean data and the true expected value of the RDE (left), and the fit of the variance of the RDE model to the variance of the data and the true variance of the RDE with $I = 100$ and a measurement noise level of $\sigma_\epsilon^2 = 1$.

I	μ_Φ	σ_Φ^2	x_0	K	σ_ϵ^2
3	2.0897	0.1762	9.4342	99.3468	1.3008
10	2.0582	0.2089	9.5680	99.9755	1.4208
30	2.0417	0.2127	9.7705	99.9530	2.1629
50	2.0450	0.2130	9.7529	99.9105	2.1755
100	2.0085	0.2004	9.9939	100.0305	2.3386
True value	2.0	0.2	10	100	4.0

TABLE 2. Estimated values for the logistic model example with measurement noise level $\sigma_\epsilon^2 = 4$. Both the simulated data and the model are assumed to have a normally distributed growth rate.

I	μ_Φ	σ_Φ^2	x_0	K	σ_ϵ^2
3	2.1649	0.1980	8.7711	99.0550	15.4414
10	2.0518	0.2144	9.7421	100.2071	22.0233
30	2.0378	0.2156	9.8703	99.9581	22.6175
50	2.0376	0.2157	9.8034	99.9554	23.0953
100	2.0047	0.2024	10.0612	100.1363	22.7998
True value	2.0	0.2	10	100	25.0

TABLE 3. Estimated values for the logistic model example with measurement noise level $\sigma_\epsilon^2 = 25$. Both the simulated data and the model are assumed to have a normally distributed growth rate.

Even in the presence of an increase in the measurement noise, our estimation procedure performs remarkable well for this example. In Tables 2 and 3 we present the results obtained with $\sigma_\epsilon^2 = 4$ and 25, respectively.

3.3. Parametric estimation - model with a misspecified distribution

Now consider the case where the growth rates are log-normally distributed, yet the distribution in which we use to model the growth rates remains a normal distribution. This is an example of model discrepancy since there is a fundamental difference between the models that are used to generate and approximate the data, however at this time we make no attempt to account for the model discrepancy in our formulation. We first consider the case where the log-normal distribution has the parameters $(\mu_\Phi, \sigma_\Phi^2) = (0.5, 0.075)$ so that the mean and variance of the true distribution are 1.7117 and 0.2282, respectively. Next we take $(\mu_\Phi, \sigma_\Phi^2) = (0.3, 0.1)$ as the parameters of the log-normal distribution so that the mean and variance are 1.4191 and 0.2118, respectively. The results of the estimated mean and variance of the distribution are presented in Table 4. We see that we obtain close approximations for the mean value and the carrying capacity; however, the variance is under approximated in both cases, more so for the second case. This is due to the fact that the log-normal distribution becomes more one sided with the choice of $(\mu_\Phi, \sigma_\Phi^2) = (0.3, 0.1)$. We expect that if we attempt to use a symmetric (in this case a normal) distribution to approximate the moments of a non-symmetric distribution (here a log-normal distribution), then our approximations will diverge from the true moments. The disparity between the approximated and true moments depends upon the degree to which the true distribution is asymmetric and, as shown above, will affect the models ability to account for variation in the data.

I	$\mathbb{E}[p_\Phi]$	$\text{Var}[p_\Phi]$	K
100	1.7363	0.1839	99.9172
True value	1.7117	0.2282	100
100	1.4177	0.1530	99.3526
True value	1.4191	0.2118	100

TABLE 4. Estimated values for the logistic model example. The data is simulated using a log-normally distributed growth rate with measurement noise $\sigma_\epsilon^2 = 1$, while the model is assumed to have a normally distributed growth rate.

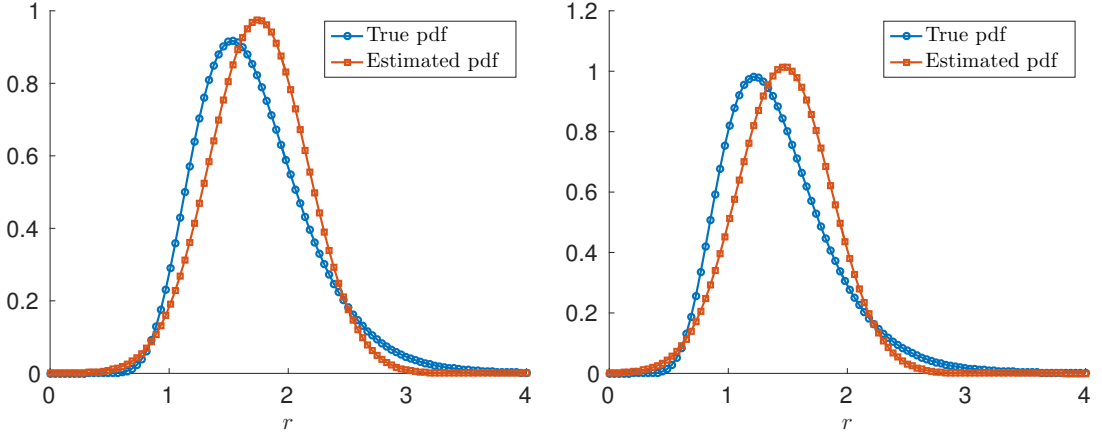


FIGURE 4. The estimated normal probability density to the true log-normal probability density with parameters $(\mu_\phi, \sigma_\phi^2) = (0.5, 0.075)$ (left) and $(\mu_\phi, \sigma_\phi^2) = (0.3, 0.1)$ (right) with $I = 100$ in both cases.

3.4. Nonparametric estimation

Using the logistic growth example, Liouville's equation gives

$$\begin{aligned} \frac{\partial u}{\partial t} + \frac{\partial}{\partial x} \left(\phi x \left(1 - \frac{x}{K} \right) u \right) &= 0 \\ u(0, x, \phi) &= u_0(x, \phi), \end{aligned} \quad (3.6)$$

which we can rewrite as

$$\begin{aligned} \frac{\partial u}{\partial t} + \phi x \left(1 - \frac{x}{K} \right) \frac{\partial u}{\partial x} &= \phi \left(\frac{2x}{K} - 1 \right) u \\ u(0, x, \phi) &= u_0(x, \phi). \end{aligned} \quad (3.7)$$

As we mentioned previously, this is a first order linear equation, and we can solve it explicitly using the method of characteristics. The characteristic curves are described by the system of ordinary differential equations

$$\begin{aligned} \frac{dt}{d\xi} &= 0, & t(\xi = 0, \nu, \rho) &= 0, \\ \frac{dx}{d\xi} &= \phi x(1 - x/K), & x(\xi = 0, \nu, \rho) &= \nu, \\ \frac{d\phi}{d\xi} &= 0, & \phi(\xi = 0, \nu, \rho) &= \rho, \\ \frac{du}{d\xi} &= \phi(2x/K - 1)u, & u(\xi = 0, \nu, \rho) &= u_0(\nu, \rho). \end{aligned}$$

Solving this system gives

$$u(\xi, \nu, \rho) = u_0(\nu, \rho) \left(\frac{\nu e^{\rho\xi} - \nu + K}{K} \right)^2 e^{-\rho\xi}, \quad (3.8)$$

where

$$\xi = t, \quad \rho = \phi, \quad \nu(t, x, \phi) = \frac{Kx}{e^{\phi t}(K - x) + x}. \quad (3.9)$$

Then, we obtain the probability density for $x(t)$ by

$$\begin{aligned} p_x(t, x) &= \int_{\Omega} u(t, \nu(t, x, \phi), \phi) d\phi \\ &= \int_{\Omega} u_0(\nu(t, x, \phi), \phi) \left(\frac{\nu(t, x, \phi)e^{\phi t} - \nu(t, x, \phi) + K}{K} \right)^2 e^{-\phi t} d\phi. \end{aligned} \quad (3.10)$$

If we assume that X_0 and Φ are independent and have probability densities p_{x_0} and p_{ϕ} , respectively, then their joint pdf is given by $u_0(x, \phi) = p_{x_0}(x)p_{\phi}(\phi)$.

For simplicity, we assume that p_{x_0} is known. What follows can readily be extended to the situation where both p_{ϕ} and p_{x_0} are unknown. However, we remark that in the case where p_{x_0} is unknown, if one has ample data collected at time $t = 0$, then the density of p_{x_0} can be estimated directly from the data by either a parametric or nonparametric approach so long as X_0 and Φ are independent.

By (2.13) we have that

$$p_{\phi} \approx p_{\phi}^M = \sum_{m=1}^M w_m \ell_m(\phi),$$

thus, we obtain

$$u_0(x, \phi) \approx u_0^M(x, \phi) = p_{x_0}(x) p_{\phi}^M(\phi) = p_{x_0}(x) \sum_{m=1}^M w_m \ell_m(\phi). \quad (3.11)$$

Substituting this expression into (3.10) gives

$$p_x(t, x) \approx p_x^M(t, x) = \int_{\Omega} p_{x_0}(\nu) \sum_{m=1}^M w_m \ell_m(\phi) \left(\frac{\nu e^{\phi t} - \nu + K}{K} \right)^2 e^{-\phi t} d\phi. \quad (3.12)$$

Given data points p_{jk} for the density of $x(t)$ at time t_j , the inverse problem is then given by

$$(\hat{\theta}, \hat{w}) = \underset{(\theta, w) \in \Theta}{\operatorname{argmin}} \mathcal{J}(\theta, w), \quad (3.13)$$

where the cost functional is now defined as

$$\mathcal{J}(\theta, w) = \sum_{k=1}^{n_x} \sum_{j=1}^{n_t} \left(p_{jk}(y_j^{(i)}) - p_x^M(t_j, x_k) \right)^2. \quad (3.14)$$

To simulate the data, the procedure is the same as described in Section 3. In this case however, the number of sampling times is set to $n_t = 51$, the initial condition was taken as $X_0 \sim \mathcal{N}(\mu_{X_0}, \sigma_{X_0}^2)$, with $\mu_{X_0} = 20$, $\sigma_{X_0}^2 = 1$, and the measurement error variance and the carrying capacity are set to $\sigma_{\epsilon} = 1$ and $K = 100$, respectively.

We first consider the case where the true probability density of Φ was taken to be normally distributed with mean and variance $\mu_{\phi} = 1$ and $\sigma_{\phi}^2 = 0.01$, respectively. In the left panel of Figure 5, we show an example of the evolution of the probability density of $x(t)$ over time. The number of approximating spline functions was set to $M = 20$, and with this we obtained a good estimation for the true density (see Figure 6), and the estimated value for the carrying capacity was $K = 99.9733$.

Next we consider the case where the probability density of Φ was that of a Bi-Gaussian distribution, which is given by

$$p_{\phi}(\phi) = \frac{1}{2\sqrt{2\pi\sigma_{\phi_1}^2}} \exp\left(-\frac{(\phi - \mu_{\phi_1})^2}{2\sigma_{\phi_1}^2}\right) + \frac{1}{2\sqrt{2\pi\sigma_{\phi_2}^2}} \exp\left(-\frac{(\phi - \mu_{\phi_2})^2}{2\sigma_{\phi_2}^2}\right),$$

with $\mu_{\phi_1} = 1.0$, $\mu_{\phi_2} = 1.5$, and $\sigma_{\phi_1}^2 = \sigma_{\phi_2}^2 = 0.005$. With this choice, the evolution of p_x over time can be seen in the right panel of Figure 5. The estimated density using $M = 30$ is given in Figure 6, and the estimate value for the carrying capacity is $K = 100.0224$, where we again obtain a good fit to the true distribution.

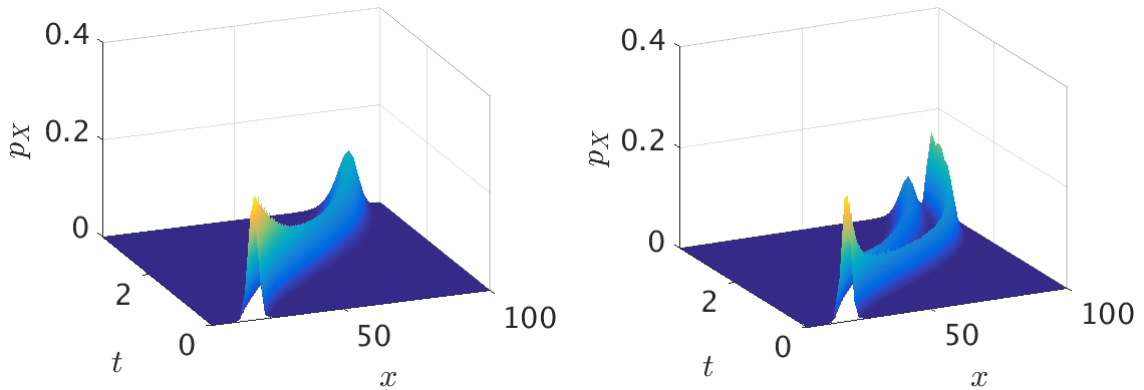


FIGURE 5. The probability density $p_X(t, x)$ obtained from the data points $y_j^{(i)}$ with noise level $\sigma_\epsilon = 1$ where p_ϕ is taken to be normal (left) and Bi-Gaussian(right).

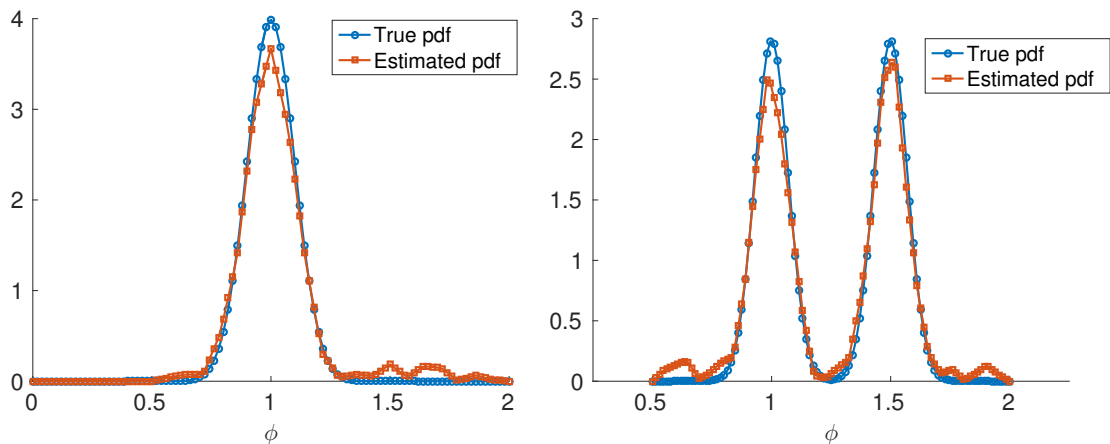


FIGURE 6. The true and estimated probability density functions. The true density is a Gaussian and the number of spline functions is $M = 20$. For the plot on the right the true density is a Bi-Gaussian and $M = 30$ spline functions were used to approximate the density.

4. Applications to a PBPK model

In this section we apply our methods to an experimental data set in which the concentration of 1,1'-Diocadecyl-3,3,3',3'-Tetramethylindotricarbocyanine Iodide dye-encapsulated nanoparticles (DIR-NPs) was collected in mice. The concentrations were measured in the liver, spleen, kidney, and plasma. The data was taken from [14] in which a PBPK model was developed specially for this experiment; the *in vivo* experiment is described fully in [14, 16]. In the experiment, at each observation time point the organs were harvested from three mice. Thus, we have three replicate measurements for the concentration of each measured compartment at each observation time point which provide an estimate of the mean and variance of the concentration over time.

We begin by modifying the original deterministic model proposed in [14] to better account for the accumulation of the drug concentration in the plasma compartment. DIR-NPs were injected into each mouse at the tail at a concentration of 5 mcg/mL and the plasma concentration measurement was collected by submandibular bleeding. In [14], the model developed by Gilkey et al. treats the injection as

a step input at $t = 0$. The authors acknowledge that this leads to a discrepancy between the model and the experimental data which exhibits a lag time in the uptake into the plasma, but justify their treatment by observing that the data and model agree after several hours have passed. We propose the following adjustment to the model to better account for the lag time in the uptake of the drug. We define a new state $D(t)$ to account for the amount of the available mass of the DIR-NPs which is taken up by the plasma. We assume that the initial concentration in the plasma is 0, and that the drug enters the plasma through a linear source term. The model accounts for the concentration of NPs in the plasma, liver, kidney, spleen and an “other” compartment. The other compartment represents the remainder of the connective pathways in the body. It is assumed that the absorption by the various tissues is thermodynamically limited rather than rate limited, resulting in first order absorption terms. The modified model is given as

$$\begin{aligned}
\frac{dC_P^{(i)}}{dt} &= \frac{1}{V_P} \left[\left(\frac{C_L^{(i)}}{R_L^{(i)}} Q_L^{(i)} + \frac{C_S^{(i)}}{R_S^{(i)}} Q_S + \frac{C_K^{(i)}}{R_K^{(i)}} Q_K + \frac{C_O^{(i)}}{R_O} Q_O \right) \right. \\
&\quad \left. - C_P^{(i)} (Q_L^{(i)} + Q_S + Q_K + Q_O) + c^{(i)} D^{(i)} \right] \\
\frac{dC_L^{(i)}}{dt} &= \frac{1}{V_L} \left[\left(C_P^{(i)} (Q_L^{(i)} - Q_S) + \frac{C_S^{(i)}}{R_S^{(i)}} Q_S + \frac{C_O^{(i)}}{R_{LO}} Q_O \right) - \frac{C_L^{(i)}}{R_L^{(i)}} (Q_L^{(i)} + Q_O) \right] \\
\frac{dC_S^{(i)}}{dt} &= \frac{1}{V_S} \left[\left(C_P^{(i)} Q_S + \frac{C_O^{(i)}}{R_{SO}^{(i)}} Q_O \right) - \frac{C_S^{(i)}}{R_S^{(i)}} (Q_S - Q_O) \right] \\
\frac{dC_K^{(i)}}{dt} &= \frac{1}{V_K} \left[\left(C_P^{(i)} Q_K + \frac{C_O^{(i)}}{R_{KO}^{(i)}} Q_O \right) - \frac{C_K^{(i)}}{R_K^{(i)}} (Q_K + Q_O + K_K^{(i)}) \right] \\
\frac{dC_O^{(i)}}{dt} &= \frac{1}{V_O} \left[Q_O \left(C_P^{(i)} - \frac{C_O^{(i)}}{R_O} \right) \right] \\
\frac{dD^{(i)}}{dt} &= -c^{(i)} D^{(i)}.
\end{aligned} \tag{4.1}$$

with $C_P^{(i)}(0) = C_L^{(i)}(0) = C_S^{(i)}(0) = C_K^{(i)}(0) = C_O^{(i)}(0) = 0, D^{(i)}(0) = D_0$, where $C_z^{(i)}$ denotes the concentration of nanoparticles in the organ z and for individual i , Q_z is the volumetric flow rate of plasma through organ z , V_z is the volume of organ z , and K_K its the clearance rate from the kidney. Due to their influence on the variance, the parameters $Q_L^{(i)}, K_K^{(i)}, R_L^{(i)}, R_S^{(i)}, R_K^{(i)}, R_{SO}^{(i)}$, and $c^{(i)}$ are taken to be the individual parameters and the remainder of the parameters are taken to be population level parameters. A log-normal distribution was chosen for each individual level parameter since each parameter must be non-negative in order to be biologically feasible.

The time varying mean and variance of the resulting RDE system (4.1) were estimated using stochastic collocation in which a sparse grid Gauss-Hermite quadrature rule was used. The mean concentration as well as the standard deviation were fit to the experimental values according to the cost function (2.12). In Figures 7–10 we show the fit of the mean and the variance of the RDE system (4.1) to the experimental data. Additionally, we compare the fit of the original deterministic model (referred to as the Gilkey model in the legend) to the mean of the experimental data. The estimated individual parameter distributions are given in Figure 11. We observe that the RDE model captures the mean data similar to the Gilkey model for $t > 5$. The modified dosing term not only provides a better fit to the initial data points ($t < 5$) but also allowed for the initial non-zero variance values in Figures 7–10. Directly estimating the distribution of the individual parameters in the RDE model allow us to obtain excellent approximations of the variance in the liver and kidney, a reasonable approximation to the spleen, and a fair fit to the plasma prior to $t = 5$. We are not able to replicate the experimental variance in the plasma after 5 hours. This may be in part due to the fact that the model itself (either the Gilkey model or our modification used here) does not conserve the mass of the administered dose of NPs. If the mass were conserved, we may be able to

better account for the variability in the plasma concentration, but adjusting the model to that degree is beyond the scope of this paper and is left for future work.

Predictive intervals for the forward propagation of uncertainty were constructed by choosing a random sample from the estimated individual parameter distributions (see Figure 11) and solving the model with these parameter values. This process was performed 300,000 times and a kernel density estimator [23] was used to compute the 95% prediction interval of the forward solutions at each time point. The predictive intervals give a maximum concentration in the Liver, Kidney, and Spleen at 300 hours of 1.5, 0.3, and 0.4 $\mu\text{g}/\text{mL}$, respectively. Each of these predictions is roughly three times higher than the value determined by the original deterministic model. The predictive intervals are representative of the variance reported in the experimental data, which can be seen by comparing the predictive intervals to the error bars of the data. The results obtained using this RDE approach yield significantly larger predictive intervals in contrast to traditional modeling PK/PD approaches where aggregate data is treated as individual data. These increased predictive intervals provide more realistic estimates of future data measurements as can be seen from the encapsulation of the error bars by the predictive intervals (see Figures 7–10 and compare Figure 3 with Figure 2). Furthermore, the increase in the uncertainty bounds on the estimated parameters can be impactful when the parameter value itself is a quantity of interest, for instance when utilizing the estimated parameter values utilizing the estimated parameter values in other mathematical models involving DIR-NP's.

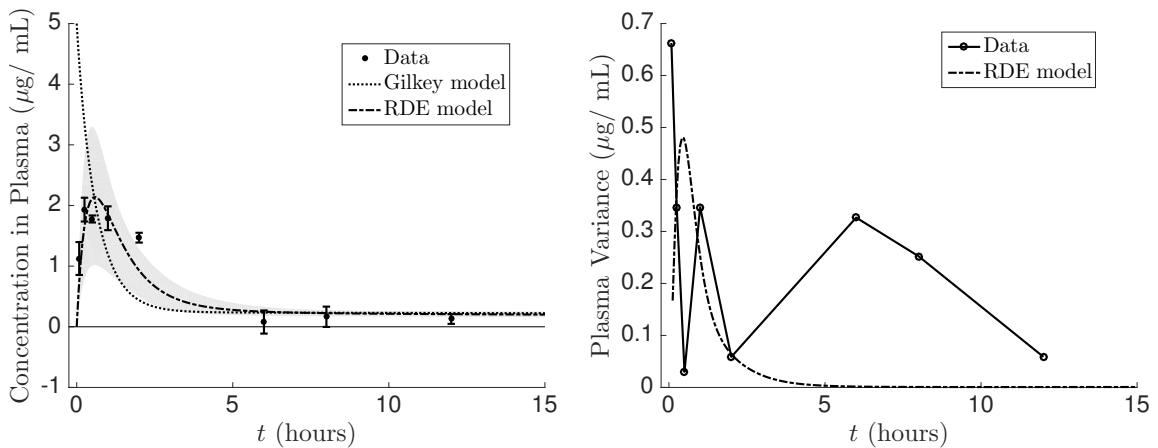


FIGURE 7. The observed concentration of nanoparticles in the plasma compared to the model fits using the original deterministic model and the expected value of the RDE system (4.1) (left). The variance computed from three replicate measurements and the fit to the computed variance of the RDE system (4.1) (right).

5. Concluding remarks

In summary, we have developed a methodology that accounts for parameter variation due to inter-individual variability for the case of repeated measurements of unidentified subjects. Obtaining repeated measurements from unidentified subjects is a common practice in many physiological experiments, and the methodology developed in this work agrees conceptually with the experimental design. Using a logistic example with synthetic data we have illustrated how this methodology can be implemented using both a parametric and nonparametric representation for the unknown individual parameter distributions. Additionally, we used experimental data obtained from repeated measurements of the concentrations of an administered dose of nano particles in a mouse to further illustrate the feasibility of the application of

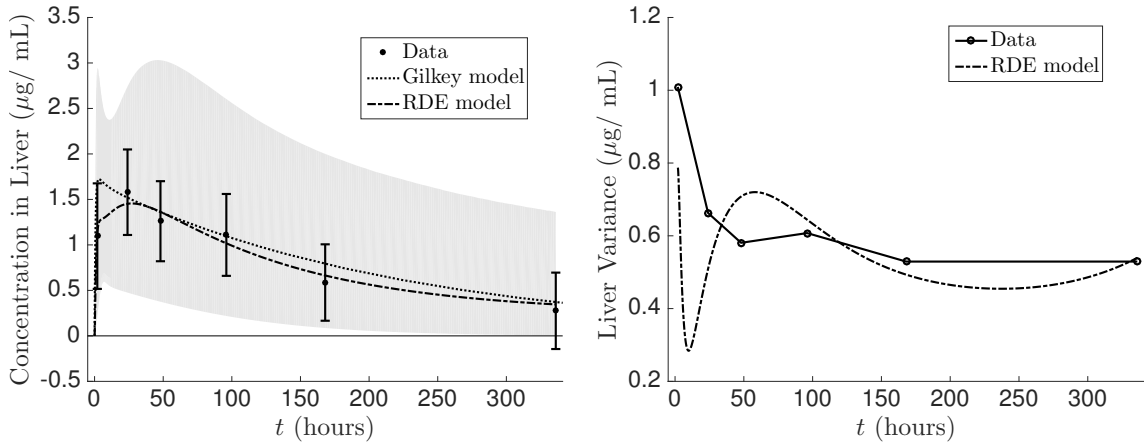


FIGURE 8. The observed concentration of nanoparticles in the liver compared to the model fits using the original deterministic model and the expected value of the RDE system (4.1) (left). The variance computed from three replicate measurements and the fit to the computed variance of the RDE system (4.1) (right).

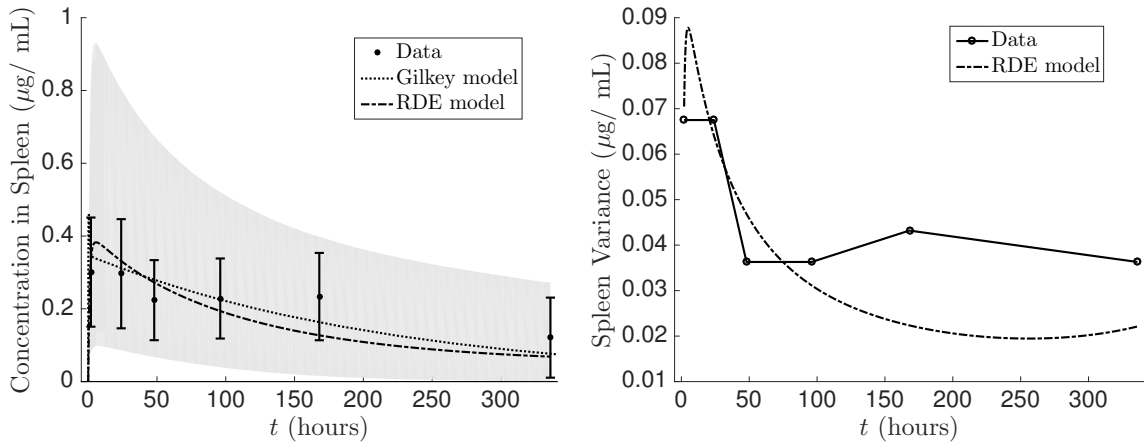


FIGURE 9. The observed concentration of nanoparticles in the spleen compared to the model fits using the original deterministic model and the expected value of the RDE system (4.1) (left). The variance computed from three replicate measurements and the fit to the computed variance of the RDE system (4.1) (right).

our methods to a recently developed PBPK model. We observed that accounting for the inter-individual variability as outlined in this work leads to differences (typically larger ranges) in the predicted quantities of interest compared to traditional methods which ignore the contribution of inter-individual variability. Recognizing and capturing this uncertainty may prove to be invaluable in the risk assessment of future experiments.

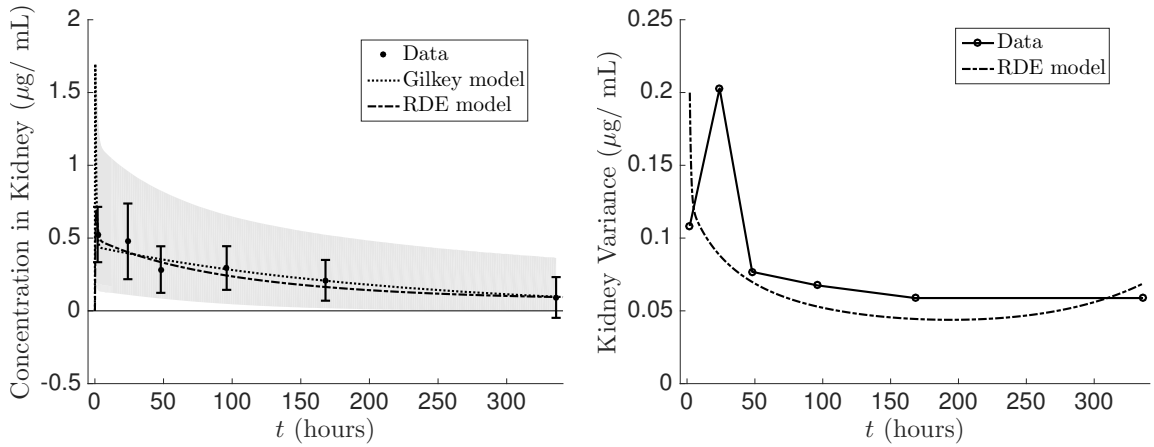


FIGURE 10. The observed concentration of nanoparticles in the kidney compared to the model fits using the original deterministic model and the expected value of the RDE system (4.1) (left). The variance computed from three replicate measurements and the fit to the computed variance of the RDE system (4.1) (right).

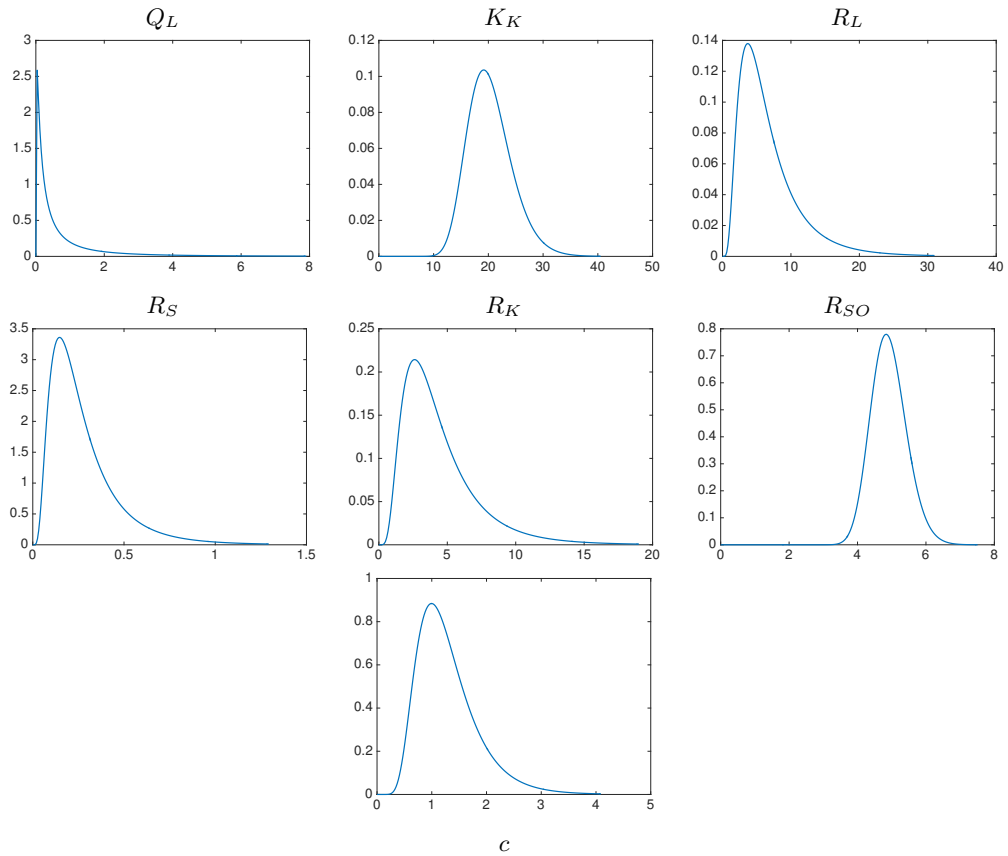


FIGURE 11. The estimated log-normal distributions for the individual parameters.

Acknowledgements. This work has been supported in part by the US Department of Education Graduate Assistance in Areas of National Need (GAANN) under grant number P200A120047 and in part by the Air Force Office of Scientific Research under grant numbers AFOSR FA9550-12-1-0188 and AFOSR FA9550-15-1-0298, and in part by the National Science Foundation under Research Training Grant (RTG) DMS- 1246991 and in part under NSF Undergraduate Biomathematics grant number DBI-1129214.

References

- [1] R.A. Albanese, H.T. Banks, M.V. Evans, L.K. Potter. *Physiologically based pharmacokinetic models for the transport of trichloroethylene in adipose tissue*. Bulletin of Mathematical Biology, 64 (2002), 97–131.
- [2] H.T. Banks. *Functional Analysis Framework for Modeling, Estimation and Control in Science and Engineering*. Chapman and Hall/CRC Press, Boca Raton, FL, 2012.
- [3] H.T. Banks, K.L. Bihari. *Modeling and estimating uncertainty in parameter estimation*. Inverse Problems, 17 (2001), 95–111.
- [4] H.T. Banks, B.G. Fitzpatrick, L.K. Potter, Y. Zhang. *Stochastic Analysis, Control, Optimization and Applications: a Volume in Honor of W.H. Fleming. Estimation of probability distributions for individual parameters using aggregate population data*. Birkhauser, Boston, 1999
- [5] H.T. Banks, S. Hu, W.C. Thompson. *Modeling and Inverse Problems in the Presence of Uncertainty*. Taylor/Francis-Chapman/Hall-CRC Press, Boca Raton, FL, 2014.
- [6] H.T. Banks, F. Kappel. *Spline approximations for functional differential equations*. Journal of Differential Equations, 34 (1979), 496–522.
- [7] H.T. Banks, Z.R. Kenz, W.C. Thompson. *A review of selected techniques in inverse problem nonparametric probability distribution estimation*. Journal of Inverse and Ill-Posed Problems, 20 (2012), 429–460.
- [8] H.T. Banks, L.K. Potter. *Probabilistic methods for addressing uncertainty and variability in biological models: Application to a toxicokinetic model*. Mathematical Biosciences, 192 (2004), 193–225.
- [9] H.A. Barton, W.A. Chiu, R.W. Setzer, M.E. Anderson, A.J. Bailer, F.Y. Bois, R.S. DeWoskin, S. Hays, G. Johanson, N. Jones, G. Loizou, R.C. MacPhail, C.J. Portier, M. Spendiff, Y.M. Tan. *Characterizing uncertainty and variability in physiologically based pharmacokinetic models: state of the science and needs for research and implementation*. Toxicological Sciences, 99 (2007), 395–402.
- [10] J.C. Caldwell, M.V. Evans, K. Krishnan. *Cutting edge PBPK models and analyses: providing the basis for future modeling efforts and bridges to emerging toxicology paradigms*. Journal of Toxicology, 2012 (2012).
- [11] M. Davidian, A.R. Gallant. *The nonlinear mixed effects model with a smooth random effects density*. Biometrika, 80 (1993), 475–488.
- [12] S. Donnet, A. Samson. *A review on estimation of stochastic differential equations for pharmacokinetic/pharmacodynamic models*. Advanced Drug Delivery Reviews, 65 (2013), 929–939.
- [13] T. Eissing, J. Lippert, S. Willmann. *Pharmacogenomics of codeine, morphine, and morphine-6-glucuronide*. Molecular Diagnosis & Therapy, 16 (2012), 43–53.
- [14] M.J. Gilkey, V. Krishnan, L. Scheetz, X. Jia, A.K. Rajasekaran, P.S. Dhurjati. *Physiologically based pharmacokinetic modeling of fluorescently labeled block copolymer nanoparticles for controlled drug delivery in leukemia therapy*. CPT: Pharmacometrics & Systems Pharmacology, 4 (2015), 167–174.
- [15] M. Krauss, R. Burghaus, J. Lippert, M. Niemi, P. Neuvonen, A. Schuppert, S. Willmann, L. Kuepfer, L. Görlitz. *Using Bayesian-PBPK modeling for assessment of inter-individual variability and subgroup stratification*. In Silico Pharmacology, 1 (2013), 1–11.
- [16] V. Krishnan, X. Xu, S.P. Barwe, X. Yang, K. Czymmek, S.A. Waldman, R.W. Mason, X. Jia, A.K. Rajasekaran. *Dexamethasone-loaded block copolymer nanoparticles induce leukemia cell death and enhance therapeutic efficacy: a novel application in pediatric nanomedicine*. Molecular Pharmaceutics, 10 (2012), 2199–2210.
- [17] M.J. Lindstrom, D.M. Bates. *Nonlinear mixed effects models for repeated measures data*. Biometrics (1990), 673–687.
- [18] J. Lippert, M. Brosch, O. Kampen, M. Meyer, H.U. Siegmund, C. Schafmayer, T. Becker, B. Laffert, L. Görlitz, S. Schreiber, P.J. Neuvonen, M. Niemi, J. Hampe, L. Kuepfer. *A mechanistic, model-based approach to safety assessment in clinical development*. CPT: Pharmacometrics & Systems Pharmacology, 1 (2012), 1–8.
- [19] G. Loizou, M. Spendiff, H.A. Barton, J. Bessems, F.Y. Bois, M.B. d’Oyvoire, H. Buist, H.J. Clewell, B. Meek, U. Gundert-Remy, G. Goerlitz, W. Schmitt. *Development of good modelling practice for physiologically based pharmacokinetic models for use in risk assessment: the first steps*. Regulatory Toxicology and Pharmacology, 50 (2008), 400–411.
- [20] J. Pinheiro, D. Bates. *Mixed-effects Models in S and S-PLUS*. Springer Science & Business Media, New York, 2006.
- [21] M.H. Schultz. *Spline Analysis*. Prentice-Hall, Englewood Cliffs, NJ, 1973.
- [22] L.B. Sheiner. *The population approach to pharmacokinetic data analysis: rationale and standard data analysis methods*. Drug Metabolism Reviews, 15 (1984), 153–171.
- [23] R.C. Smith. *Uncertainty Quantification: Theory, Implementation and Application*. SIAM, Philadelphia, 2013.
- [24] T.T. Soong. *Random Differential Equations in Science and Engineering*. Academic Press, New York, 1973.
- [25] D. Xiu. *Numerical Methods for Stochastic Computations: A Spectral Method Approach*. Princeton University Press, Princeton, NJ, 2010.