THE PROHOROV METRIC FRAMEWORK AND AGGREGATE DATA INVERSE PROBLEMS FOR RANDOM PDEs

H.T. BANKS[1], K.B. FLORES[2], I.G. ROSEN[3], E.M. RUTTER[4],
MELIKE SIRLANCI[5], AND W. CLAYTON THOMPSON[6]

[1,2,4,6]Center for Research in Scientific Computation
Department of Mathematics
North Carolina State University
Raleigh, NC, USA

[3,5]Department of Mathematics
University of Southern California
Los Angeles, CA, USA

[6]SAS Institute
Cary, NC, USA

ABSTRACT: We consider nonparametric estimation of probability measures for parameters in problems where only aggregate (population level) data are available. We summarize an existing computational method for the estimation problem which has been developed over the past several decades [24, 5, 12, 28, 16]. Theoretical results are presented which establish the existence and consistency of very general (ordinary, generalized and other) least squares estimates and estimators for the measure estimation problem with specific application to random PDEs.

1. INTRODUCTION

Our ideas of using distributions for parameters in inverse problems grew out of work in [7] and use of Sinko-Streifer models for mosquitofish growth rates where only aggregate data were available due to periodic sampling of different subsets of the population. The first theoretical results in the context of inverse problems was in [12]. There were numerous subsequent uses including shrimp population size models in [4, 11] and carboxyfluorescein succinimidyl ester (CFSE) labeling models [13]. In our presentation here we consider approximation methods in estimation or inverse problems but the quantity of interest is a probability distribution. Assume we have a parameterized system ($q \in \Omega$) with state model responses $x(t; q)$ describing the population of interest. For data or observations, we are given a set of values $\{y_l \approx Cx(t_l; q)\}$ for the expected values

$$\mathcal{E}[y_l(q)|P] = \int_\Omega y_l(q) dP(q)$$

for observations $y_l(q) = Cx(t_l, q)$ with respect to the unknown probability distribution P describing the distribution of parameters q over the population.

We use data to choose from a given family of distributions $\mathcal{P}(Q)$ the distribution P^* that gives the best fit of the underlying model to data. This is accomplished by formulating an ordinary least squares (OLS) problem; however we note that we could equally well use a weighted least squares (WLS) or maximum likelihood estimation (MLE) framework. Specifically we seek to minimize

$$J(P) = \sum_l |\mathcal{E}[y_l(q)|P] - y_l|^2$$

over $P \in \mathcal{P}(\Omega)$. Even for simple dynamics for $y_l(q)$ this yields an infinite dimensional optimization problem. Therefore one needs approximations that lead to computationally tractable schemes. That is, it is useful to formulate methods to yield finite dimensional sets $\mathcal{P}^M(\Omega)$ over which to minimize $J(P)$. Of course, we wish to choose these methods so that "$\mathcal{P}^M(\Omega) \to \mathcal{P}(\Omega)$" in some sense. In our case we shall use the *Prohorov Metric* [5, 24] of weak star convergence of measures to assure the desired approximation convergences as well as consistency of the estimators.

The data $\{y_l\}$ available (which, in general, will involve longitudinal or time

evolution data) determines the nature of the problem. In gerneral there are three classes of problems:

- **Type I**: The most classical problem (which we shall refer to as a *Type I* problem) is one in which individual longitudinal data is available for each member in the population. In this case there is a wide statistical literature (in the context of hierarchical modeling, mixing distributions, mixed or random effects, mixture models, etc.) [29, 33, 34, 35, 36, 50, 49, 51, 53, 65, 64, 63] which provides theory and methodology for estimating not only individual parameters but also population level parameters and allows one to investigate both intra-individual and inter-individual variability in the population and data.

- **Type II**: In what we shall refer to as *Type II* problems one has only *aggregate* or population level longitudinal data available. This is common in marine, insect, etc., *catch and release* experiments [17] where one samples at different times from the same population but cannot be guaranteed of observing the same subset of individuals at each sample time under constant environmental, etc., conditions. This type of data is also typical in experiments where the organism or population member being studied is sacrificed in the process of making a single observation (e.g., certain physiologically based pharmacokinetic (PBPK) modeling [1, 21, 66, 57] and whole organism transport models [17]). In this case one may still have dynamic (i.e., time course) models for individuals, but no individual data are available.

- **Type III**: Finally, the third class of problems which we shall refer to as *Type III* problems involves dynamics which depend explicitly on the probability distribution P itself. In this case one only has dynamics (*aggregate dynamics*) for the expected value

$$\bar{x}(t) = \mathcal{E}[x(t, q, P)|P]$$

of the state variable. No dynamics are available for individual trajectories $x(t, q)$ for a given $q \in Q$. Such problems arise in *viscoelasticity* and *electromagnetics* as well as biology (the HIV cellular models of [6]) see also [5, 15, 14, 20, 44].

While the approximations we discuss below are applicable to all three types of problems, our primary interest here is problems of Type II and Type III. In particular we shall illustrate and provide theoretical underpinnings for our earlier computational results in the context of *Transdermal Alcohol Concentration (TAC)* and *Glioblastoma Multiforme (GBM) cancer* where the inverse problems are of Type II. Finally, we note that in the problems considered here, one *can not sample directly from the probability distribution* being estimated and this again is somewhat different from the usual case treated in some of the statistical literature, e.g., see [81, 82] and the references cited therein.

We first describe two relevant examples currently under investigation: the first is a linear partial differential equation with unknown diffusion; the second is a nonlinear diffusion-growth system for tumors.

2. EXAMPLE 1: ESTIMATING BLOOD/BREATH ALCOHOL CONCENTRATION FROM TRANSDERMAL ALCOHOL CONCENTRATION(TAC)

The measurement of the alcohol level in the human body for the purpose of medical research, clinical therapy, and law enforcement (e.g. DUI, etc.), typically takes the form of blood alcohol concentration (BAC). However, in the absence of a blood sample, which is almost always the case, a surrogate, breath alcohol concentration (BrAC) as measured by an instrument known as a breath analyzer, is used. The underlying chemistry of the breath analyzer (based on Henry's Law [48]) has been shown to be reasonably robust and consistent across individuals and ambient conditions, thus allowing for the relatively straight forward conversion of breath alcohol concentration (BrAC) to blood alcohol concentration (BAC). Unfortunately, however, there are two significant drawbacks to collecting data using a breath analyzer: properly blowing into a breath analyzer so as to obtain accurate measurements can be challenging, and the collection of breath data that is near-continuous in time and in a naturalistic setting is all but impossible.

Recently, technology has been developed to allow for the measurement of transdermal alcohol, or alcohol that diffuses from the skin's dermal layer which has an active blood supply, through the epidermal layer of the skin. These de-

vices use a variety of technologies (electro-chemical, enzymatic, optical, etc.) to count the number of ethanol molecules evaporating from the surface of the skin through normal perspiration. The current generation of transdermal alcohol biosensors are reasonably compact and relatively unobtrusive, and generally resemble a digital watch, ankle bracelet, or activity tracker. Two such devices are the WrisTAS^{TM}7 alcohol biosensor designed and manufactured by Giner, Inc. of Waltham, MA and the Secure Continuous Alcohol Monitoring System (SCRAM) device manufactured by Alcohol Monitoring Systems in Littleton, Colorado (see Figure 1). These devices offer the possibility of pas-

Figure 1: Alcohol Biosensor Devices: The WrisTAS (left) and the SCRAM (right).

sively collecting drinking and intoxication data both in the lab or clinic and out in the field, that is essentially continuous in time over extended periods such as hours, days, or even weeks. It is also conceivable that they could be further miniaturized and included as a feature in the next generation of wearable health monitoring technology. At present, however, with the exception of one company that monitors abstinence of DUI offenders under contract to the courts, these devices are primarily only being used in the research community, with the devices themselves (their utility, practicality, accuracy, dependability, etc.) the central focus of the research project. This is because while it has been known for a long time that the alcohol level in perspiration correlates well with the alcohol concentration in the blood or breath [77, 78, 79, 80], there are significant variations (1) from sensor to sensor and (2) in the rate at which alcohol diffuses through the skin both across individuals and across distinct drinking episodes within individuals under differing environmental conditions. Consequently the meaningful quantitative interpretation of transdermal alcohol levels poses a significant challenge. More to the point, there is currently

no known direct and generally accepted method for converting what these devices measure, transdermal alcohol concentration or TAC, to the quantities that researchers, clinicians, law enforcement and the public at large are all most familiar with and that are well understood measures of intoxication, BAC and/or BrAC.

Some of the difficulties involved in converting TAC to BAC or BrAC are illustrated in the following two example data sets. A WrisTAS^{TM}7 was worn by a participant for 18 days. During each drinking episode, the participant collected BrAC data (i.e., blew into a breath analyzer) approximately every 30 minutes. The first drinking episode was conducted in the laboratory and BrAC was measured every 15 minutes until it returned to 0.000. The participant then wore the device for the following 17 days and consumed alcohol ad libitum. During those days, BrAC was measured every 30 minutes starting from the beginning of the drinking session until its value returned to 0.000. The WrisTAS^{TM}7 measured and recorded ethanol level at the skin surface every 5-minutes. During those 17 days, the data were collected in a naturalistic setting. The plot in Figure 2 shows the measured BrAC and TAC over 11
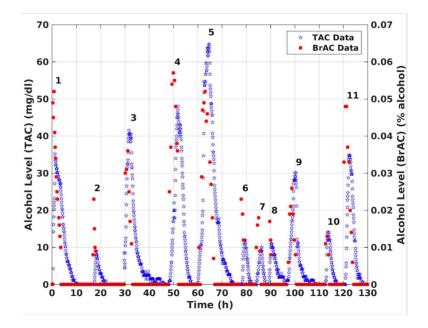

Figure 2: BrAC and TAC measurements for multiple drinking episodes by a single subject.

drinking episodes. Note that in drinking episodes 1, 2, 4, 6, 7, 8, and 11 the peak BrAC value was higher than the bench calibrated peak TAC value, while in drinking episodes 3, 5, 9, and 10 the peak TAC value was higher than the peak BrAC value.

In a second example, data were collected from multiple subjects at the University of Illinois at Urbana-Champaign using the AMS SCRAM alcohol biosensor. In this study 60 participants were given a gender and weight adjusted dose of alcohol calculated to yield a peak BrAC of approximately .08%. The participants were then asked to provide breath analyzer readings at 30 minute intervals. The SCRAM sensor was worn and provided transdermal readings also at a rate of one every 30 minutes. We then fit the two parameters in the first principles physics based forward model to be described below to 18 of the 60 data sets collected. While for each participant's data, the resulting models produced estimated BrAC that closely fit the observed BrAC (obtained by inverting the forward model) and estimated TAC (obtained by putting the estimated BrAC back through the forward model) that essentially reproduced the measured or observed TAC, as the scatter plot below clearly shows, there was nevertheless a wide variance in the values of the fit parameters.

A number of researchers have looked at the TAC to BAC/BrAC conversion problem (see, for example, [37, 38, 39]) with the goal being to produce reliable quantitative estimates of BrAC/BAC (eBrAC/eBAC) from TAC data. In a series of papers [32, 40, 59, 68] one of the co-authors of this paper (IGR) together with other members of his research group devised a two-step protocol that used individual calibration data (i.e., simultaneously-collected breath analyzer BrAC measurements and biosensor TAC measurements) to first fit a first principles physics/physiological-based forward model to capture the transport of ethanol from the blood, through the skin, and its eventual processing and measurement by the TAC sensor. They then used the fit forward model together with deconvolution techniques to estimate the eBrAC/eBAC input to the model from the TAC output for other drinking episodes not included in the training set. Such an approach has the obvious drawback of requiring that the forward model be tuned to each individual and device and the ambient environmental conditions (e.g., temperature, humidity, etc.). Indeed, our algorithms have been used in alcohol related consumption and behavioral
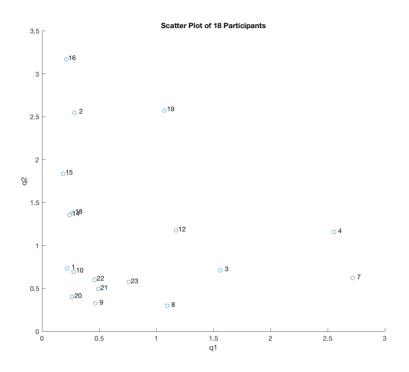
Figure 3: Scatter plot of deterministically obtained parameter esti-
mates for BrAC and TAC measurements from a single drinking episode
from each of 18 subjects.

studies [52] in which drinking patterns in individuals with and without alco-
hol metabolizing genetic variants were investigated, and in [41] to investigate
the relationship between social familiarity and alcohol reward in naturalistic
drinking settings and compared this to alcohol reward observed in laboratory
drinking settings. The results of these studies clearly indicated that at least
a portion of the dynamics of the system are not being captured by the mod-
els. In addition, the requisite calibration procedure is a significant burden to
researchers, clinicians and participants. They may also pose a number of eth-
ical concerns (e.g., requiring the offering of alcoholic beverages to participants
known to be pregnant or suffering from alcohol use disorder).

In a more recent series of papers [69, 67, 70], this same research group has
proposed eliminating the need to calibrate the forward models, by developing
population models based on their physics based models. The parameters in

the model are assumed to be random and instead of using their actual values as a basis for fitting the model, the training data is used to estimate the distribution of these random parameters. In those studies, however, the arguments used to demonstrate the convergence of their method required the assumption that the distributions were described by appropriately constrained parameterized families of probability density functions. The theory we describe here demonstrates how this restrictive assumption can be eliminated.

By essentially converting all quantities in our first principles physics based model to dimensionless variables, we are able to eliminate some non-identifiable (i.e., dependent) parameters and obtain the initial-boundary value problem given by

$$\frac{\partial \varphi}{\partial t}(t, \eta) = q_1 \frac{\partial^2 \varphi}{\partial \eta^2}(t, \eta), \quad 0 < \eta < 1, \ t > 0, \tag{1}$$

$$q_1 \frac{\partial \varphi}{\partial \eta}(t, 0) = \varphi(t, 0), \quad t > 0, \tag{2}$$

$$q_1 \frac{\partial \varphi}{\partial \eta}(t, 1) = q_2 u(t), \quad t > 0, \tag{3}$$

$$\varphi(0, \eta) = \varphi_0, \quad 0 < \eta < 1, \tag{4}$$

$$y(t) = \varphi(t, 0), \quad t > 0, \tag{5}$$

where the partial differential equation (1) describes the one dimensional transport of ethanol through the epidermal layer of the skin, and the boundary conditions (2) and (3) describe respectively the evaporation of ethanol at the skin surface ($\eta = 0$) and the flux of ethanol across the boundary between the epidermal and dermal layers of the skin ($\eta = 1$; note that the dermal layer has a blood supply whereas the epidermal layer does not). The initial conditions are given in (4) (typically we will have $\varphi_0 = 0$ since we will assume that at the start of any drinking episode there is no alcohol in the body) and the output equation, (5) says that TAC is measured by the biosensor at the skin surface. The state of the system, $\varphi(t, \eta)$ essentially denotes the concentration of ethanol at depth $\eta, 0 \leq \eta \leq 1$ in the epidermal layer at time $t \geq 0$, while $u(t)$ and $y(t)$ denote, respectively, the BAC/BrAC and TAC at time t. The parameters q_1 (normalized "diffusion") and q_2 (normalized flux control input coefficient) are assumed to lie in a compact subset, Q, of $\mathbb{R}^+ \times \mathbb{R}^+$ endowed with the metric d_Q.

While our ultimate goal is to deconvolve the BAC/BrAC input u from the biosensor measured TAC output y, our concern in this study is dealing with the uncertainty in the model parameters. Indeed, since the parameters q which are unknown and dependent on 1) the individual wearing the sensor, 2) the ambient environmental conditions at the time the TAC measurements or observations were made, and 3) the particular sensor being worn, the problem of estimating BAC/BrAC from observations of TAC is in fact a blind deconvolution problem involving aggregate data. Our approach involves replacing the q-dependent model with a population model which captures and quantifies the variability and uncertainty that exists across all the individual members of the population (we use the term *individual* to mean not just different subjects, but also *unmodelled different environmental conditions and different hardware*, and for that matter *any other un-modeled dynamics present in the system*). Thus we view our problem as the class of Type II described in the Introduction. The details involved in deconvolving the BAC/BrAC from the TAC once the population model has been estimated along with associated error bars can be found in [68].

We define the *population model* to be the system (1)-(5) with the parameters $\mathbf{q} \in Q$ defined to be a random vector together with the distribution for the parameters in the form of a probability measure P_0 or joint distribution function F_0. We assume (1) that ν data sets $(\tilde{u}_i, \tilde{y}_i)_{i=1}^{\nu} = \left(\{\tilde{u}_{i,j}\}_{j=0}^{\nu_i-1}, \{\tilde{y}_{i,j}\}_{j=0}^{\nu_i} \right)_{i=1}^{\nu}$ have been collected, and (2) the statistical model given by

$$\tilde{Y}_{i,j} = \mathbb{E}[Y_j(\mathbf{q}, u_i)|P_0] + \mathcal{E}_{i,j}, \quad j = 0, ..., \nu_i, \ i = 1, ..., \nu, \tag{6}$$

where in (6), $Y_j(\mathbf{q}, u_i)$ is the random variable corresponding to (1)-(5) and $\mathcal{E}_{i,j}, \ j = 0, ..., \nu_i, \ i = 1, ..., \nu$, represent measurement noise and are assumed to be independent and identically distributed with mean 0 and common variance σ^2. Then for P a probability measure defined on the σ-algebra of events in the probability space associated with the random vector $\mathbf{q}$, and F its corresponding cumulative distribution function, define for an observation operator $\hat{C}(q)$ the the mean behavior at time j, $j = 0, ..., \nu_i$,

$$v_i(j; P) = \mathbb{E}[y_j(q, u_i)|P] = \int_Q \hat{C}(q) x_j(q, u_i) dF(q), \tag{7}$$

where $q \sim F$ and y_j is a realization for Y_j .

The estimation problem is then to estimate the measure P_0, using a least squares approach

$$\hat{P}_0 = \operatorname*{argmin}_P J(P; (\tilde{u}_i, \tilde{y}_i)_{i=1}^{\nu}) = \operatorname*{argmin}_P \sum_{i=1}^{\nu} \sum_{j=0}^{\nu_i} (\tilde{y}_{i,j} - v_i(j; P))^2. \qquad (8)$$

where the $v_i(j; P)$ are as given in (7).

3. EXAMPLE 2: GLIOBLASTOMA MULTIFORME (GBM)

Glioblastoma Multiforme (GBM) is a deadly primary brain tumor. Due to it's highly infiltrative nature, GBM remains difficult to treat: although resection surgery may remove the primary tumor, many tumor cells can remain throughout the brain, resulting in nearly all tumors recurring [83]. It has been shown that tumors that exhibit nodular growth patterns (i.e., low diffusivity) result in better patient prognosis [2]. Thus, it is increasingly important to be able to accurately ascertain the growth and diffusion phenotypes present in individual tumors.

GBM is often modeled *in vivo* using partial differential equations. The simplest models of GBM growth only consider the reaction-diffusion equation [55, 75, 61], given by:

$$\frac{\partial c(t, x)}{\partial t} = D \frac{\partial^2 c(t, x)}{\partial x^2} + \rho c(t, x)(1 - c(t, x)) \qquad (9)$$

where $c(t, x)$ represents the cell density at time t and spatial location x, D is the diffusion coefficient, and ρ is the intrinsic growth rate. However, this simplistic model assumes diffusion is isotropic, which does not accurately describe resulting *in vivo* tumor shapes.

There are multiple methods to incorporate anisotropic diffusion in order to make glioma models more physiologically accurate. These methods introduce heterogeneity into the diffusion coefficients. For example, one can incorporate a spatially-dependent piece-wise diffusion coefficient to explain diffusion of cells in grey matter versus white matter [76]. In a similar vein, diffusion weighted imaging (DWI), which contains information about water diffusion, has been used to infer cellular diffusion through the brain [30, 31, 46, 47, 56]. Other

approaches to incorporating cellular heterogeneity include density-dependent diffusion functions [73].

A landmark study discovered that, even *in vitro*, a reaction-diffusion equation was insufficient to describe the cellular dynamics [71]. In the work, the authors performed cell line experiments on most common mutation of Epidermal Growth Factor Receptor gene (U87ΔEGFR) and wild-type EGFR (U87WT). This showed there are distinct behavioral differences between 'migrating' cells and 'proliferating cells' and thus concluded that migrating and proliferating cells should be modeled separately. This observation resulted in the modeling of cellular heterogeneity as a collection of two phenotypic subpopulations, one of which mainly migrated and one that mainly proliferated. This "go or grow" hypothesis [45] remains widely used in mathematical models of glioma growth today [72, 43]. For a recent review on the status of mathematical modeling in GBM see [54].

We proposed in [60] to perform an inverse problem to estimate the diffusion D, and growth rate ρ in the reaction-diffusion equation as distributions of a random differential equation rather than as point estimates for a deterministic differential equation. In this case any spatiotemporal data involves a growing/decreasing population of different individual cells and even if one has frequent measurements of the changing populations, one expects (because of a continuum of growth/death of the individual cells) this to represent (as in the case of many biological cell data counts [13]) aggregates of different populations at different times of data collections.

In [60], we proposed a model of GBM growth with the option of phenotypic heterogeneity by using parameter distributions for the parameters ρ and D. The random differential equation for GBM growth and diffusion is given by:

$$\frac{\partial c(t,x;\boldsymbol{D},\boldsymbol{\rho})}{\partial t} = \nabla \cdot (\boldsymbol{D}\nabla c(t,x;\boldsymbol{D},\boldsymbol{\rho})) + \boldsymbol{\rho}c(t,x;\boldsymbol{D},\boldsymbol{\rho})(1 - c(t,x;\boldsymbol{D},\boldsymbol{\rho})) \quad (10)$$

where $\boldsymbol{D}$ and $\boldsymbol{\rho}$ are random variables defined on a compact set $\Omega = \Omega_{\boldsymbol{D}} \times \Omega_{\boldsymbol{\rho}}$. In order to obtain the aggregate observable, we take the expectation over all subpopulations:

$$v(t,x) = \mathbb{E}\left[c(t,x;\cdot,\cdot),P\right] = \int_{\Omega} c(t,x;\boldsymbol{D},\boldsymbol{\rho})dP(\boldsymbol{D},\boldsymbol{\rho}) \quad (11)$$

The given random differential equation has the option of phenotypic heterogeneity because we can easily recover the basic reaction-diffusion equation

(Eq. 9) when $\boldsymbol{D}$ and $\boldsymbol{\rho}$ are point distributions. Moreover, we are able to model a "go or grow" system by considering bigaussian distributions of the parameters. Most importantly, we are able to model these distributions (and more) without making any assumptions on the underlying distributions.

Several different methods for approximating the probability measures $P(\boldsymbol{D}, \boldsymbol{\rho})$ include using either discrete approximations based on *delta functions* or continuous approximations based on *spline basis functions*. Although using higher order spline functions are known [3, 9, 10, 20] to yield more accurate convergence in the case of smooth probability density functions (PDF) and cumulative distribution functions (CDF), delta functions are able to better approximate CDFs that have discontinuous derivatives. Therefore, in [60] we discussed the use of both approximations since we did not wish to make any assumptions about, or restrictions on, the CDF.

Suppose that the aggregate spatiotemporal data we want to model is given by v_{ji}, representing the data at time j and spatial location i, where $j = 1, .., N_t$ and $i = 1, ..., N_x$. Then, we estimate using $N = N_t \times N_x$ data points

$$\hat{P}_N = \underset{\mathcal{P}(\Omega)}{\operatorname{argmin}} \sum_{j,i=1}^{N_t, N_x} (v_{ji} - v(t_j, x_i; P))^2.$$

(This, of course, is an idealized version of available data sets but we will use this notation when discussing consistency. We assume in our subsequent discussions involving $N \to \infty$ statements it is understood both $N_t, N_x \to \infty$.)

This becomes

$$\hat{P}_N = \underset{\mathcal{P}(\Omega)}{\operatorname{argmin}} \sum_{j,i} \left(v_{ji} - \int_{\Omega} c(t_j, x_i; \boldsymbol{D}, \boldsymbol{\rho}) dP(\boldsymbol{D}, \boldsymbol{\rho}) \right)^2 \tag{12}$$

where c is the solution of the random differential equation (10).

4. PROHOROV ESTIMATES AND THEIR APPROXIMATIONS

Convergence in the Prohorov metric is equivalent to the weak* convergence of measures when the space of probability measures $\mathcal{P}(\Omega)$ is imbedded in the dual $C^*(\Omega)$ of the space of bounded continuous functions on Ω. We discuss briefly existence, convergence, and consistency theory, assuming here we are

only estimating $\boldsymbol{D}$ of the GBM example above as a distribution. We note the theory readily extends to two parameters (diffusion and growth rates) for the GBM example and to the vector parameters in the case of the TAC example discussed above.

In the following discussions we adopt the following notation : $N = (N_t, N_x) = N_t \times N_x$ are the number of data points or observations; $v^{\tilde{N}}(t, x; \omega)$ are the state approximations to $v(t, x; \omega)$, so $\tilde{N}$ is the index for state approximations; M will be the parameter approximation index

We assume a family $\mathcal{P}(\Omega)$ of permissible probability functions for our diffusion rates. We attempt to perform the estimation in a least-squares framework

$$\hat{P}_N = \operatorname*{argmin}_{P \in \mathcal{P}(\Omega)} J_N(\vec{v}, P) = \operatorname*{argmin}_{P \in \mathcal{P}(\Omega)} \sum_{j,i}^{N} (v_{ji} - v(t_j, x_i; P))^2 \tag{13}$$

where N observations are used to obtain a best fit for a nominal or "true" parameter P_0. In order to approximate this minimizer, we replace the infinite dimensional optimization problem by a sequence of finite-dimensional optimization problems with, for example, Dirac or spline-based distributions. Thus, if we use the Dirac approximating families, we set $\boldsymbol{D}^M = \{\Delta_{D_k}, k = 1, .., M\}$, where M represents the number of nodes, or elements, used in the approximation. Our family of approximating probability functions becomes

$$\mathcal{P}^M(\Omega) = \left\{ P^M = \sum_{k=1}^{M} w_k \Delta_{D_k} | w_k \geq 0 \text{ and } \sum_{k=1}^{M} w_k = 1 \right\},$$

where Δ_{D_k} represents the Dirac delta function at the point D_k and w_k are the weights and/or probabilities. It has been previously proven [23, 28] that there exists a minimizer for the discrete approximation problem

$$\hat{P}_N^M = \operatorname*{argmin}_{P \in \mathcal{P}^M(\Omega)} \sum_{j,i=1}^{N_t, N_x} (v_{ji} - v(t_j, x_i; P))^2. \tag{14}$$

A further approximation arises when we appproximate the state variable v by numerical approximation $v^{\tilde{N}}$, e.g., by finite elements for example, and seek to solve

$$\hat{P}_{N,M}^{\tilde{N}} = \operatorname*{argmin}_{P \in \mathcal{P}^M(\Omega)} \sum_{j,i=1}^{N_t, N_x} (v_{ji} - v^{\tilde{N}}(t_j, x_i; P))^2. \tag{15}$$

There are a number of questions that arise immediately in the class of problems we have defined. Perhaps the most obvious are questions of convergence (what happens as $M \to \infty$ in the Dirac or spline approximations?) and consistency (what happens as $N = (N_t, N_x) \to \infty$?) These questions have been successfully investigated both theoretically ([28, 20, 22] and the references therein) and computationally ([10, 9] and the references therein) for certain classes of problems. A further issue involves the partial differential equation approximations $c_{\tilde{N}}$ (e.g., finite element approximations of the realizations of the the random PDE (10)) to the solution c. Again, elements of the necessary convergence issues have been addressed in [28, 17]. In summary we wish to establish for the problems discussed here that the approximations $\hat{P}_{N,M}^{\tilde{N}}$ converge to a "true" distribution P_0 as the number of elements used in the approximations increases (i.e., $M, N, \tilde{N} \to \infty$).

Significantly, the Prohorov Metric Framework is computationally constructive. That is, in practice, one does not construct a sequence of estimates for increasing values of M and $\tilde{N}$; rather, one fixes the values of M and $\tilde{N}$ to be sufficiently large to attain a desired level of accuracy. To do this we need only to have some enumeration of the elements of $\mathcal{P}^M(\Omega)$ in order to compute an approximate estimate $\hat{P}_{N,M}^{\tilde{N}}$. Practically, this is accomplished by selecting M nodes in Ω, $\{\omega_k\}_{k=1}^M$. The optimization problem (15) is then reduced to a standard constrained estimation problem over Euclidean M-space in which one determines the values of the weights p_k^M corresponding to each node. Thus,

$$\hat{P}_{N,M}^{\tilde{N}} = \arg \min_{\mathcal{P}^M(\Omega)} \sum_{j,i} \left(v_{ji} - v^{\tilde{N}}(t_j, x_i; P) \right)^2 \tag{16}$$

$$= \arg \min_{\mathcal{P}^M(\Omega)} \sum_{j,i} \left(v_{ji} - \int_\Omega c_{\tilde{N}}(t_j, x_i; \omega) dP(\omega) \right)^2 \tag{17}$$

$$= \arg \min_{\widetilde{\mathbb{R}^M}} \sum_{j,i} \left(v_{ji} - \left(\sum_{k=1}^M c_{\tilde{N}}(t_j, x_i; \omega_k) p_k^M \right) \right)^2, \tag{18}$$

where in the final line we seek the weights $\bar{p}^M = (p_1^M, \ldots, p_M^M)^T \in \widetilde{\mathbb{R}^M} = \{\bar{p}^M | p_k^M \in \mathbb{R}^+, \sum_{k=1}^M p_k^M = 1\}$. These are sufficient to characterize the approximating discrete estimate $\hat{P}_{N,M}^{\tilde{N}}$ since the $\tilde{N}$ nodes are assumed to be

fixed in advance. Moreover, define

$$H_{kl} = 2 \sum_{j,i} \left(c_{\tilde{N}}(t_j, x_i; \omega_k) \right) \left(c_{\tilde{N}}(t_j, x_i; \omega_l) \right)$$

$$f_k = -2 \sum_{j,i} v_{ji} \left(c_{\tilde{N}}(t_j, x_i; \omega_k) \right)$$

$$d = \sum_{j,i} (v_{ji})^2 .$$

Then one can equivalently compute [10]

$$\hat{P}_{N,M}^{\tilde{N}} = \arg \min_{\mathbb{R}^{\widetilde{M}}} \left(\frac{1}{2} \left(\bar{p}^M \right)^T H \bar{p}^M + f^T \bar{p}^M + d \right). \tag{19}$$

From this reformulation, it is clear that the approximate problem (16) has a unique solution if H is positive definite and the minimum occurs in the interior of the space. If the individual mathematical model is independent of P (See [23, Sec. 14.1.2] for a complete discussion) then the matrices H and f can be precomputed in advance. Then one can rapidly (and exactly) compute the gradient and Hessian of the objective function in a numerical optimization routine. As M grows large, the quadratic optimization problem (19) becomes poorly conditioned [10, 9] and there exists a trade-off: M must be chosen sufficiently large so that the computational approximation is accurate, but not so large that ill-conditioning leads to large numerical errors. The efficient choice of M as well as the choice of the nodes $\{\omega_k\}_{k=1}^M$ are open research problem-dependent questions.

It should be acknowledged that the uniqueness of the computational problem (i.e., when H is positive definite) is not sufficient to ensure the uniqueness of the limiting estimate $\hat{P}_N^*$ in Theorem 5.4 below (as there could be multiple convergent subsequences). However, if $J_N(\vec{v}; P)$ of (13) is uniquely minimized, then every subsequence of $\hat{P}_{N,M}^{\tilde{N}}$ which converges must converge to that unique minimizer. Moreover, under assumptions (A1)–(A7) detailed below, it can be shown that $\frac{1}{N} J_N(\vec{v}, P) \to J_0(P)$ (as N grows large) with probability one, and the function $J_0(P)$ is assumed to be uniquely minimized by P_0.

To illustrate the ideas, we continue our discussions with $\Omega = \Omega_{\boldsymbol{D}}$ taken as the continuum of values in $[0, D_{\max}]$. Hence the family of probability functions $\mathcal{P}(\Omega)$ is a compact metric space in the Prohorov metric. The minimizer function is continuous in P; and there exists a (not necessarily unique) minimizer $\hat{P}$.

5. RESULTS FOR RANDOM PDE MODELS

We first note that the results of [22] are given in terms of estimators P_N and estimates $\hat{P}_N$ for nonlinear *random ordinary differential equations*

$$\frac{dy}{dt} = g(t, y(t); \boldsymbol{q}) \tag{20}$$

$$y(t_0) = y_0. \tag{21}$$

We claim that the results of [22] hold immediately if we replace the random DEs with the random partial differential equations (10) or (1)-(5). This can be readily established by a careful reading of all the details of [22]. We summarize and discuss the resulting RDE details.

5.1. Existence of the Estimator

For RDE models one can then prove the existence [22, Thm 3.1] of P_N and $\hat{P}_N$ as measurable functions mapping a subset of $\mathbb{R}^N$ (that is, the data $\vec{v} \in \mathbb{R}^N$ where $N = (N_t, N_x)$ in the case of our GBM example) into the space of probability measures on Ω. We remark that the statement of the existence theorem concerns the estimate $\hat{P}_N$ obtained from the data realizations $\vec{v} \in \mathbb{R}^N$ of the random vector $\vec{V}$. This is sufficient to establish the existence of the estimator P_N as a measurable function as well, since the random vector $\vec{V}$ is by definition a measurable function from a probability triple into $\mathbb{R}^N$, and the composition of measurable functions is measurable. We note that the nonlinearities such as those in the GBM example discussed above or the vector nature of the parameters in the TAC example above play *no role* in the proofs given in [22]. We therefore restate the existence results here without proof, referring the reader to [22] for further details.

Theorem 5.1. Define the function $J_N : \mathbb{R}^N \times \mathcal{P}(\Omega) \to \mathbb{R}$ according to Equation (13). Assume (Ω, d) is separable and compact and take the space of probability measures $\mathcal{P}(\Omega)$ with the Prohorov metric ρ. Assume further that $J_N(\cdot, P)$ is a measurable function from $\mathbb{R}^N \to \mathbb{R}$ for each $P \in \mathcal{P}(\Omega)$, and that $J_N(\vec{v}, \cdot) : \mathcal{P}(\Omega) \to \mathbb{R}$ is continuous for each $\vec{v} \in \mathbb{R}^N$. Then there exists a

measurable function $\hat{P}_N : \mathbb{R}^N \to \mathcal{P}(\Omega)$ such that

$$J(\vec{v}, \hat{P}_N(\vec{v})) = \inf_{P \in \mathcal{P}(\Omega)} J(\vec{v}, P).$$

5.2. Consistency of the Estimator

We can next establish consistency for estimators in the case of random partial differential equation problems. The assumptions are essentially the same as those in the case of ordinary differential equation estimators.

(A1) For any fixed $N = N_t \times N_x$, the error random variables $\{\mathcal{E}_j\}_{j=1}^N$ are independent and identically distributed, defined on some probability triple $(\Theta, \Sigma_\Theta, P_\Theta)$.

(A2) For $\vec{\mathcal{E}} = (\mathcal{E}_1, \dots, \mathcal{E}_N)$, $E[\vec{\mathcal{E}}] = 0$ and $Cov[\vec{\mathcal{E}}] = \sigma^2 I_N$, where I_N is the $N \times N$ identity matrix.

(A3) (Ω, d) is a separable, compact metric space; the space $\mathcal{P}(\Omega)$ is taken with the Prohorov metric ρ.

(A4) For all j, $1 \le j \le N_t$, i, $1 \le i \le N_x$, $(t_j, x_i) \in \tilde{T}$ for some compact space $\tilde{T}$.

(A5) The model function $v \in C(\mathcal{P}(\Omega), C(\tilde{T}))$.

(A6) There exists a measure μ on $\tilde{T}$ such that for all $g \in C(\tilde{T})$

$$\frac{1}{N} \sum_{j,i=1} g(t_j, x_i) \equiv \int_{\tilde{T}} g(t, x) d\mu_N(t, x) \to \int_{\tilde{T}} g(t, x) d\mu(t, x)$$

(A7) The functional $J_0(P) = \int_{\tilde{T}} (v(t, x; P_0) - v(t, x; P))^2 d\mu(t, x)$ is uniquely minimized at $P_0 \in \mathcal{P}(\Omega)$.

Under the assumptions one can prove consistency.

Theorem 5.2. Under assumptions (A1)-(A7), there exists a set $A \in \Sigma_\Theta$ with $P_\Theta(A) = 1$ such that for all $\theta \in A$,

$$\frac{1}{N} J_N(\vec{V}; P)(\theta) \to J_0(P)$$

as $N \to \infty$ and for each $P \in \mathcal{P}(\Omega)$. Moreover, the convergence is uniform on $\mathcal{P}(\Omega)$.

Theorem 5.3. Under assumptions (A1)-(A7), the estimators $P_N \xrightarrow{w^*} P_0$ as $N \to \infty$ with probability 1. That is,

$$P_\Theta \left(\left\{ \theta \Big| P_N(\vec{V})(\theta) \to P_0 \right\} \right) = 1.$$

Complete proofs of these two theorems are given in [22].

Theorem 5.3 establishes the consistency of the estimator (12). Given a set of data $\vec{v}$, it follows that the estimate $\hat{P}_N$ corresponding to the estimator P_N will converge to the true distribution P_0 under the stated assumptions. We remark that these assumptions are not overly restrictive (compare [12, 28, 42]) though some of the assumptions may be difficult to verify in practice. Assumptions (A3)–(A5) are mathematical in nature and may be verified directly for each specific problem. Assumptions (A1) and (A2) describe the error process which is assumed to generate the collected data. While it may be possible to ascertain a priori that the error process satisfies these assumptions (see [8]), one may also use posterior analysis such as residual plots [27, Ch. 3] to investigate the appropriateness of the assumptions of the statistical model. Assumption (A6) reflects the manner in which data is sampled and, together with Assumption (A7), constitutes an identifiability condition for the model. The limiting sampling distribution function μ may be known if the experimenter has complete control over the values t_j, x_i of the independent variables (e.g., if the t_j, x_i are measurement times and locations) but this may not always be the case.

The novel results in [22] establishes the desirable property of *consistency of the estimator* P_N as a measurable function mapping the data observation process to the space of probability measures. However, it is generally not possible to directly solve the optimization problems (13) for $\hat{P}_N$ as a function of $\vec{v}$. As a result, approximate (generally numerical) methods must be used in order to solve (15) and obtain an approximate estimate $\hat{P}_{N,M}^{\tilde{N}}$. We must ascertain, then, how the approximate estimate $\hat{P}_{N,M}^{\tilde{N}}$ relates to the exact estimate $\hat{P}_N$ (for any fixed value of N.)

The following result establishes the *computational convergence* of the Prohorov Metric framework for fixed N. These results establish a comprehensive body of theory for the least squares estimation of the measure P_0 that is assumed to have generated the observed data. It is given in [22, Theorem 5.1]:

Theorem 5.4. (Convergence) Let (Ω, d) be a compact, separable metric space and consider the space $(\mathcal{P}(\Omega), \rho)$ of probability measures on Ω with the Prohorov metric, as before. Let $\mathcal{P}^M(\Omega)$ be as defined as above (e.g., using Dirac or spline approximates for elements of Ω). Assume

1. The map $P \mapsto J_N^{\tilde{N}}(\vec{v}, P)$ is continuous for all $\tilde{N}, N$;

2. For any sequence of probability measures $P_k \to P$ in $\mathcal{P}(\Omega)$, $v^{\tilde{N}}(t, x; P_k) \to v(t.x; P)$ as $\tilde{N}, k \to \infty$;

3. The function $v(t, x; P)$ is uniformly bounded for all t, x, P.

Then there exists minimizers $\hat{P}_{N,M}^{\tilde{N}}$ satisfying (15). Moreover, for fixed N, there exists a subsequence (as $M, \tilde{N} \to \infty$) of the approximate estimates $\hat{P}_{N,M}^{\tilde{N}}$ which converges to some $\hat{P}_N^*$ which satisfies (13).

This theorem provides a set of conditions under which a sequence of approximate estimates $\hat{P}_{N,M}^{\tilde{N}}$ converges to the estimate $\hat{P}_N$ of interest. This estimate is itself a realization (for a particular data set) of the estimator P_N which has been shown to exist and to be consistent, so that $P_N \to P_0$ with probability one. Thus we are assured that a computed measure $\hat{P}_{N,M}^{\tilde{N}}$ is an accurate estimate of the true distribution P_0. The assumptions of Theorem 5.4 are not restrictive. In typical problems (and, indeed, in the assumptions of other theorems appearing in this document) it is assumed that the parameter space Ω as well as the independent variable space $T \times X$ are compact. In such a case, Assumptions 1 and 3 above are satisfied if the individual model solutions $c(t, x; \boldsymbol{D}, \boldsymbol{\rho})$ are continuous on $T \times X \times \Omega$. Assumption 2 is then simply a condition on the convergence of the numerical procedures used in obtaining model solutions which we turn to discuss now! This result is, in essence, a verification of an analogue for our problems of hypothesis (iv) of Theorem 3.1 of [12].

6. ESTIMATION AND APPROXIMATION ARGUMENTS

Recall that we are interested in minimizing for fixed N the least squares cost functional (22):

$$J_N(P) = \sum_{j,i}^{N} |v_{ji} - v(t_j, x_i; P)|^2.$$
(22)

We first argue the following:

Theorem 6.1. Suppose $v^{\tilde{N}}(t, x; \omega) \to v(t, x; \omega)$ for each t, x, uniformly in ω in Ω, and that, for each t, x, the mapping $\omega \to v(t, x; \omega)$ is continuous on Ω. Suppose $P_k \to P$ in $\mathcal{P}(\Omega)$. Then for each t, x as $\tilde{N}, k \to \infty$ we have

$$v^{\tilde{N}}(t, x; P_k) = \int_\Omega v^{\tilde{N}}(t, x; \omega) dP_k(\omega) \to v(t, x; P) = \int_\Omega v(t, x; \omega) dP(\omega).$$

Proof. For each fixed t, x we have

$$\begin{aligned}
& \left| \int_\Omega v^{\tilde{N}}(t, x; \omega) dP_k(\omega) - \int_\Omega v(t, x; \omega) dP(\omega) \right| \\
& \leq \left| \int_\Omega \left(v^{\tilde{N}}(t, x; \omega) - v(t, x; \omega) \right) dP_k(\omega) \right| \\
& \quad + \left| \int_\Omega v(t, x; \omega) dP_k(\omega) - \int_\Omega v(t, x; \omega) dP(\omega) \right| \\
& \equiv I + II.
\end{aligned}$$
(23)

For the first term, we find

$$I \leq \int_\Omega |v^{\tilde{N}}(t, x; \omega) - v(t, x; \omega)| dP_k(\omega).$$

Suppose $\epsilon > 0$. Choose N_0 so that $\tilde{N} \geq N_0$ implies

$$\left| v^{\tilde{N}}(t, x; \omega) - v(t, x; \omega) \right| < \epsilon \quad \text{for all } \omega \in \Omega.$$

Then for every k we have

$$\int_\Omega \left| v^{\tilde{N}}(t, x; \omega) - v(t, x; \omega) \right| dP_k(\omega) < \epsilon.$$

Thus $I \to 0$ as $\tilde{N} \to \infty$ uniformly in k.

Considering the second term, we have

$$II = \left[\int_\Omega v(t, x; \omega) dP_k(\omega) - \int_\Omega v(t, x; \omega) dP(\omega) \right]$$

But for each t, x, $v(t, x; \cdot)$ is in $C(\Omega)$ and by definition of the Prohorov metric (actually, one of it's equivalent characterizations!), we have immediately that $II \to 0$ and the theorem is proved.

It remains to verify that "$v^{\tilde{N}}(t, x; \omega) \to v(t, x; \omega)$ for each t, x, uniformly in ω in Ω". But this is simply a finite element type approximation for the systems (1)-(5) or (9) [62, 74, 58]. These arguments have been given for Sinko-Streifer types hyperbolic systems [26] as well the nonlinear parabolic models of interest here [25, 17, 18, 3]. Indeed, combining the nonlinear approximation arguments of [19, p. 585; take $q^n \equiv q$ throughout] with the linear system arguments for $u^{\tilde{N}}(t, x; \omega) - \pi_{\tilde{N}} u(t, x; \omega)$ in [18, 3] yields the desired approximation results. (Here $\pi_{\tilde{N}} u(t, x; \omega)$ is the orthogonal projection on to the finite element approximation subspaces.)

7. CONCLUDING REMARKS

Inter-individual or intra-individual heterogeneity is often ignored in mathematical models. In the above discussions we model heterogeneity using random differential equation models. That is, we formulate (partial) differential equations in which some parameters are random variables. In particular, we are concerned with the ability to recover the parameter distributions without making any assumptions about the probability distributions.

We have illustrated and validated the computational results in the context of *Transdermal Alcohol Concentration (TAC)* models and *Glioblastoma Multiforme (GBM) cancer* where the inverse problems are of the aggregate data/individual model type. We have established existence of estimators in classes of probability distributions, convergence of approximations and consistency of the estimators.

ACKNOWLEDGEMENT

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 **64** (2002), no. 1, 97.

[2] A. L. Baldock, S. Ahn, R. Rockne, S. Johnston, M. Neal, D. Corwin, K. Clark-Swanson, G. Sterin, A. D. Trister, H. Malone, et al., *Patient-specific metrics of invasiveness reveal significant prognostic benefit of resection in a predictable subset of gliomas*, PloS One **9** (2014), no. 10, e99057.

[3] H. T. Banks, J. Crowley, and K. Kunisch, *Cubic spline approximation techniques for parameter estimation in distributed systems*, IEEE Transactions on Automatic Control **28** (1983), no. 7, 773–786.

[4] H. T. Banks, V. Bokil, S. Hu, A. K. Dhar, R. Bullis, C. Browdy, and F. Allmutt, *Modeling shrimp biomass and viral infection for production of biological countermeasures*, Mathematical Biosciences and Engineering **3** (2006), no. 4, 635–660.

[5] H. T. Banks, D. Bortz, G. Pinter, and L. Potter, *Modeling and imaging techniques with potential for application in bioterrorism*, Bioterrorism:

Mathematical Modeling Applications in Homeland Security, SIAM, 2003, pp. 129–154.

[6] H. T. Banks, D. Bortz, and S. Holte, *Incorporation of variability into the modeling of viral delays in hiv infection dynamics*, Mathematical Biosciences **183** (2003), no. 1, 63–91.

[7] H. T. Banks, L. W. Botsford, F. Kappel, and C. Wang, *Modeling and estimation in size structured population models*, LCDS-CCS Report 87-13, Brown University; Proc. 2nd Course on Mathematical Ecology, (Trieste, December 8–12, 1986) (1988), 521–541.

[8] H. T. Banks, J. Catenacci and S. Hu, *Use of difference-based methods to explore statistical and mathematical model discrepancy in inverse problems*, CRSC-TR15-05, May, 2015; J. Inverse Ill-Posed Problems, **24** (2016), 413–433.

[9] H. T. Banks and J. L. Davis, *Quantifying uncertainty in the estimation of probability distributions with confidence bands*, CRSC-TR07-21, December, 2007; Mathematical Biosciences and Engineering **5** (2008), 647–667.

[10] H. T. Banks and J. L. Davis, *A comparison of approximation methods for the estimation of probability distributions on parameters*, Applied Numerical Mathematics **57** (2007), no. 5-7, 753–777.

[11] H. T. Banks, J. L. Davis, S. L. Ernstberger, S. Hu, E. Artimovich, and A. K. Dhar, *Experimental design and estimation of growth rate distributions in size-structured shrimp populations*, Inverse Problems **25** (2009), no. 9, 095003.

[12] H. T. Banks and B. G. Fitzpatrick, *Estimation of growth rate distributions in size structured population models*, Quarterly of Applied Mathematics **49** (1991), no. 2, 215–235.

[13] H. T. Banks, K. B. Flores, C. R. Langlois, T. R. Serio, and S. S. Sindi, *Estimating the rate of prion aggregate amplification in yeast with a generation and structured population model*, Inverse Problems in Science and Engineering (2017), 1–23.

[14] H. T. Banks and N. L. Gibson, *Electromagnetic inverse problems involving distributions of dielectric mechanisms and parameters*, CRSC-TR05-29, August, 2005; Quarterly of Applied Mathematics **64** (2006), 749–795.

[15] H. T. Banks and N. L. Gibson, *Well-posedness in maxwell systems with distributions of polarization relaxation parameters*, Applied Mathematics Letters **18** (2005), no. 4, 423–430.

[16] H. T. Banks, Z. R. Kenz, and W. C. Thompson, *A review of selected techniques in inverse problem nonparametric probability distribution estimation*, CRSC-TR12-13, May 2012; J. Inverse and Ill-Posed Problems (2012), 429–460.

[17] H. T. Banks and K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhausen, Boston, 1989.

[18] H. T. Banks and P. Lamm, *Estimation of variable coefficients in parabolic distributed systems*, IEEE Transactions on Automatic Control **30** (1985), no. 4, 386–398.

[19] H. T. Banks and K. Murphy, *Estimation of nonlinearities in parabolic models for growth, predation, and dispersal of populations*, Journal of Mathematical Analysis and Applications **141** (1989), no. 2, 580–602.

[20] H. T. Banks and G. A. Pinter, *A probabilistic multiscale approach to hysteresis in shear wave propagation in biotissue*, Multiscale Modeling & Simulation **3** (2005), no. 2, 395–412.

[21] 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; Mathematical Biosciences **192** (2004), 193–22.

[22] H. T. Banks and W. C. Thompson, *Existence and consistency of a nonparametric estimator of probability measures in the Prohorov Metric Framework*, International Journal of Pure and Applied Mathematics **103** (2015), no. 4, 819–843.

[23] H. T. Banks, *A functional analysis framework for modeling, estimation and control in science and engineering*, CRC Press, 2012.

[24] H. T. Banks and K. L. Bihari, *Modelling and estimating uncertainty in parameter estimation*, Inverse Problems **17** (2001), no. 1, 95.

[25] H. T. Banks, P. M. Kareiva, and K. A. Murphy, *Parameter estimation techniques for interaction and redistribution models: a predator-prey example*, Oecologia **74** (1987), no. 3, 356–362.

[26] H. T. Banks and K. A. Murphy, *Quantitative modeling of growth and dispersal in population models*, Mathematical topics in population biology, morphogenesis and neurosciences, Springer, 1987, pp. 98–109.

[27] H. T. Banks and H. T. Tran, *Mathematical and experimental modeling of physical and biological processes*, CRC Press, 2009.

[28] H. T. Banks, S. Hu, and W. C. Thompson, *Modeling and inverse problems in the presence of uncertainty*, CRC Press, 2014.

[29] S. L. Beal and L. B. Sheiner, *Estimating population kinetics.*, Critical Reviews in Biomedical Engineering **8** (1982), no. 3, 195–222.

[30] P.-Y. Bondiau, O. Clatz, M. Sermesant, P.-Y. Marcy, H. Delingette, M. Frenay, and N. Ayache, *Biocomputing: numerical simulation of glioblastoma growth using diffusion tensor imaging*, Physics in Medicine and Biology **53** (2008), no. 4, 879.

[31] O. Clatz, M. Sermesant, P.-Y. Bondiau, H. Delingette, S. K. Warfield, G. Malandain, and N. Ayache, *Realistic simulation of the 3-d growth of brain tumors in mr images coupling diffusion with biomechanical deformation*, IEEE Transactions on Medical Imaging **24** (2005), no. 10, 1334–1346.

[32] Z. Dai, I. Rosen, C. Wang, N. Barnett, and S. E. Luczak, *Using drinking data and pharmacokinetic modeling to calibrate transport model and blind deconvolution based data analysis software for transdermal alcohol biosensors.*, Mathematical Biosciences and Engineering **13** (2016), no. 5, 911–934.

[33] M. Davidian and A. R. Gallant, *Smooth nonparametric maximum likelihood estimation for population pharmacokinetics, with application to quinidine*, Journal of Pharmacokinetics and Biopharmaceutics **20** (1992), no. 5, 529–556.

[34] M. Davidian and A. R. Gallant, *The nonlinear mixed effects model with a smooth random effects density*, Biometrika **80** (1993), no. 3, 475–488.

[35] M. Davidian and D. M. Giltinan, *Nonlinear models for repeated measurement data*, Chapman & Hall, 1995.

[36] M. Davidian and D. M. Giltinan, *Nonlinear models for repeated measurement data: an overview and update*, Journal of Agricultural, Biological, and Environmental Statistics **8** (2003), no. 4, 387.

[37] D. M. Dougherty, N. E. Charles, A. Acheson, S. John, R. M. Furr, and N. Hill-Kapturczak, *Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking.*, Experimental and Clinical Psychopharmacology **20** (2012), no. 5, 373.

[38] D. M. Dougherty, N. Hill-Kapturczak, Y. Liang, T. E. Karns, S. E. Cates, S. L. Lake, J. Mullen, and J. D. Roache, *Use of continuous transdermal alcohol monitoring during a contingency management procedure to reduce excessive alcohol use*, Drug & Alcohol Dependence **142** (2014), 301–306.

[39] D. M. Dougherty, T. E. Karns, J. Mullen, Y. Liang, S. L. Lake, J. D. Roache, and N. Hill-Kapturczak, *Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking*, Drug & Alcohol Dependence **148** (2015), 77–84.

[40] M. A. Dumett, I. Rosen, J. Sabat, A. Shaman, L. Tempelman, C. Wang, and R. Swift, *Deconvolving an estimate of breath measured blood alcohol concentration from biosensor collected transdermal ethanol data*, Applied Mathematics and Computation **196** (2008), no. 2, 724–743.

[41] C. E. Fairbairn, K. Bresin, D. Kang, I. G. Rosen, T. Ariss, S. E. Luczak, N. P. Barnett, and N. S. Eckland, *A multimodal investigation of contextual effects on alcohol's emotional rewards.*, Journal of Abnormal Psychology **127** (2018), no. 4, 359.

[42] A. R. Gallant, *Nonlinear statistical models*, John Wiley & Sons, 1987.

[43] P. Gerlee and S. Nelander, *Travelling wave analysis of a mathematical model of glioblastoma growth*, Mathematical Biosciences and Engineering **276** (2016), 75–81.

[44] N. L. Gibson, *Terahertz-based electromagnetic interrogation techniques for damage detection*, Ph.D. thesis, North Carolina State University, August 2004.

[45] H. Hatzikirou, D. Basanta, M. Simon, K. Schaller, and A. Deutsch, *'Go or grow': the key to the emergence of invasion in tumour progression?*, Mathematical Medicine and Biology **29** (2012), no. 1, 49–65.

[46] S. Jbabdi, E. Mandonnet, H. Duffau, L. Capelle, K. R. Swanson, M. Pélégrini-Issac, R. Guillevin, and H. Benali, *Simulation of anisotropic growth of low-grade gliomas using diffusion tensor imaging*, Magnetic Resonance in Medicine **54** (2005), no. 3, 616–624.

[47] E. Konukoglu, O. Clatz, P.-Y. Bondiau, M. Sermesant, H. Delingette, and N. Ayache, *Towards an identification of tumor growth parameters from time series of images*, Medical Image Computing and Computer-Assisted Intervention–MICCAI 2007, Springer, 2007, pp. 549–556.

[48] D. A. Labianca, *The chemical basis of the breathalyzer: A critical analysis*, Journal of Chemical Education **67** (1990), no. 3, 259.

[49] B. G. Lindsay, *Mixture models: theory, geometry and applications*, NSF-CBMS regional conference series in probability and statistics, JSTOR, 1995, pp. i–163.

[50] B. G. Lindsay et al., *The geometry of mixture likelihoods: a general theory*, The Annals of Statistics **11** (1983), no. 1, 86–94.

[51] B. G. Lindsay and M. L. Lesperance, *A review of semiparametric mixture models*, Journal of Statistical Planning and Inference **47** (1995), no. 1-2, 29–39.

[52] S. E. Luczak, I. G. Rosen, and T. L. Wall, *Development of a real-time repeated-measures assessment protocol to capture change over the course of a drinking episode*, Alcohol and Alcoholism **50** (2015), no. 2, 180–187.

[53] A. Mallet, *A maximum likelihood estimation method for random coefficient regression models*, Biometrika **73** (1986), no. 3, 645–656.

[54] N. L. Martirosyan, E. M. Rutter, W. L. Ramey, E. J. Kostelich, Y. Kuang, and M. C. Preul, *Mathematically modeling the biological properties of gliomas: A review*, Mathematical Biosciences and Engineering **12** (2015), no. 4, 879–905.

[55] J. D. Murray, *Mathematical biology ii: Spatial models and biomedical applications*, 3 ed., Springer-Verlag New York Incorporated, 2003.

[56] K. Painter and T. Hillen, *Mathematical modelling of glioma growth: the use of diffusion tensor imaging (dti) data to predict the anisotropic pathways of cancer invasion*, Journal of Theoretical Biology **323** (2013), 25–39.

[57] L. K. Potter, *Physiologically based pharmacokinetic models for the systemic transport of trichloroethylene*, Ph.D. thesis, North Carolina State Univeristy, August 2001.

[58] P. M. Prenter, *Splines and variational methods*, John Wiley & Sons, 1975.

[59] I. G. Rosen, S. E. Luczak, and J. Weiss, *Blind deconvolution for distributed parameter systems with unbounded input and output and determining blood alcohol concentration from transdermal biosensor data*, Applied Mathematics and Computation **231** (2014), 357–376.

[60] E. M. Rutter, H. T. Banks, and K. B. Flores, *Estimating intratumoral heterogeneity from spatiotemporal data*, Journal of Mathematical Biology (2018).

[61] E. M. Rutter, T. L. Stepien, B. J. Anderies, J. D. Plasencia, E. C. Woolf, A. C. Scheck, G. H. Turner, Q. Liu, D. Frakes, V. Kodibagkar, Y. Kuang, M. C. Preul, and E. J. Kostelich, *Mathematical analysis of glioma growth in a murine model*, Scientific Reports **7** (2017), 2508.

[62] M. H. Schultz, *Spline analysis*, Prentice-Hall, 1973.

[63] A. Schumitzky, *The nonparametric maximum likelihood approach to pharmacokinetic population analysis*, Proceedings of the 1993 Western Simulation Multiconference: Simulation for Health Care. San Diego Society for Computer Simulation, 1993, pp. 95–100.

[64] A. Schumitzky, *Nonparametric em algorithms for estimating prior distributions*, Applied Mathematics and Computation **45** (1991), no. 2, 143–157.

[65] L. B. Sheiner, B. Rosenberg, and K. L. Melmon, *Modelling of individual pharmacokinetics for computer-aided drug dosage*, Computers and Biomedical Research **5** (1972), no. 5, 441–459.

[66] J. Simmons, W. Boyes, P. Bushnell, J. Raymer, T. Limsakun, A. McDonald, Y. Sey, and M. Evans, *A physiologically based pharmacokinetic model for trichloroethylene in the male long-evans rat*, Toxicological Sciences **69** (2002), no. 1, 3–15.

[67] M. Sirlanci, S. Luczak, and I. Rosen, *Approximation and convergence in the estimation of random parameters in linear holomorphic semigroups generated by regularly dissipative operators*, American Control Conference (ACC), 2017, IEEE, 2017, pp. 3171–3176.

[68] M. Sirlanci, S. E. Luczak, C. E. Fairbairn, K. Bresin, D. Kang, and I. Rosen, *Deconvolving the input to random abstract parabolic systems: a population model-based approach to estimating blood/breath alcohol concentration from transdermal alcohol biosensor data*, Submitted for publication (2018).

[69] M. Sirlanci, S. E. Luczak, C. E. Fairbairn, D. Kang, R. Pan, X. Yu, and I. Rosen, *Estimation of the distribution of random parameters in a diffusion equation forward model for a transdermal alcohol biosensor*, Submitted for publication (2017).

[70] M. Sirlanci and I. Rosen, *Estimation of the distribution of random parameters in discrete time abstract parabolic systems with unbounded input and output: Approximation and convergence*, Submitted for publication (2017).

[71] A. M. Stein, T. Demuth, D. Mobley, M. Berens, and L. M. Sander, *A mathematical model of glioblastoma tumor spheroid invasion in a three-dimensional in vitro experiment*, Biophysical Journal **92** (2007), no. 1, 356–365.

[72] T. L. Stepien, E. M. Rutter, and Y. Kuang, *Traveling waves of a go-or-grow model of glioma growth*, SIAM Journal of Applied Mathematics (2018).

[73] T. L. Stepien, E. M. Rutter, and Y. Kuang, *A data-motivated density-dependent diffusion model of in vitro glioblastoma growth*, Mathematical Biosciences and Engineering **12** (2015), no. 6, 1157–1172.

[74] G. Strang and G. J. Fix, *An analysis of the finite element method*, vol. 212, Prentice-hall Englewood Cliffs, NJ, 1973.

[75] K. Swanson, E. Alvord, and J. Murray, *A quantitative model for differential motility of gliomas in grey and white matter*, Cell Proliferation **33** (2000), no. 5, 317–330.

[76] K. R. Swanson, C. Bridge, J. Murray, and E. C. Alvord, *Virtual and real brain tumors: using mathematical modeling to quantify glioma growth and invasion*, Journal of the Neurological Sciences **216** (2003), no. 1, 1–10.

[77] R. M. Swift, *Transdermal alcohol measurement for estimation of blood alcohol concentration*, Alcoholism: Clinical and Experimental Research **24** (2000), no. 4, 422–423.

[78] R. M. Swift, *Direct measurement of alcohol and its metabolites*, Addiction **98** (2003), no. s2, 73–80.

[79] R. M. Swift, *Transdermal measurement of alcohol consumption*, Addiction **88** (1993), no. 8, 1037–1039.

[80] R. M. Swift, C. S. Martin, L. Swette, A. Laconti, and N. Kackley, *Studies on a wearable, electronic, transdermal alcohol sensor*, Alcoholism: Clinical and Experimental Research **16** (1992), no. 4, 721–725.

[81] G. Wahba, *Bayesian" confidence intervals" for the cross-validated smoothing spline*, Journal of the Royal Statistical Society. Series B (Methodological) (1983), 133–150.

[82] G. Wahba, *Splines in nonparametric regression*, Wiley Online Library, 2000.

[83] P. Y. Wen and S. Kesari, *Malignant gliomas in adults*, New England Journal of Medicine **359** (2008), no. 5, 492–507.