Automatica 106 (2019) 101–109

Contents lists available at ScienceDirect

Automatica

journal homepage: www.elsevier.com/locate/automatica

Brief paper

# Estimating the distribution of random parameters in a diffusion equation forward model for a transdermal alcohol biosensor☆

Melike Sirlanci [a], Susan E. Luczak [b], Catharine E. Fairbairn [c], Dahyeon Kang [c], Ruoxi Pan [d], Xin Yu [d], I. Gary Rosen [d,*]

[a] Department of Computing and Mathematical Sciences, California Institute of Technology, Pasadena, CA 91125, USA
[b] Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA
[c] Department of Psychology, University of Illinois Urbana-Champaign, Champaign, IL 61820, USA
[d] Modeling and Simulation Laboratory, Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA

ARTICLE INFO

*Article history:*
Received 27 October 2017
Received in revised form 31 December 2018
Accepted 31 March 2019
Available online xxxx

*Keywords:*
Distribution estimation
Biosensor data
Distributed parameter systems
Random parameters
Blood alcohol concentration
Transdermal alcohol concentration

ABSTRACT

We estimate the distribution of random parameters in a distributed parameter model with unbounded input and output for the transdermal transport of ethanol. The underlying model is a diffusion equation with input: blood alcohol concentration and output: transdermal alcohol concentration. We reformulate the dynamical system so that the random parameters are treated as additional space variables. When the distribution to be estimated is absolutely continuous with a joint density, estimating the distribution is equivalent to estimating the diffusivity in a multi-dimensional diffusion equation. Well-established finite dimensional approximation schemes, functional analytic based convergence arguments, optimization techniques, and computational methods may be employed. We use our technique to estimate a bivariate normal distribution based on data for multiple drinking episodes from a single subject.



## 1. Introduction

Researchers and clinicians studying and treating alcohol dependence have long sought the means to continuously and quantitatively monitor blood alcohol levels in naturalistic settings. The ability to do this would be extremely valuable for advancing a wide range of alcohol research and clinical treatment domains, including how alcohol concentrations relate to drinking motives, physical responses to alcohol, behaviors, decision-making, negative consequences, and environmental situational factors over the course of a drinking episode and across drinking episodes. At present the only ways to collect alcohol consumption data in the field are the self-report diary and the breath analyzer, both of which require the active participation of the subject and generally yield inaccurate estimates of blood alcohol level when obtained during naturalistic drinking episodes.

Recently, biosensors that measure transdermal alcohol concentration (TAC), the amount of alcohol diffusing through the skin (Fig. 1), have been developed and used, but primarily only as abstinence monitors (for example, in court mandated monitoring of DUI offenders). Unfortunately, however, the transport and filtering of alcohol by the skin is affected by a number of factors that differ across individuals (e.g., skin layer thickness, porosity, and tortuosity, etc.) and across drinking episodes within individuals (e.g., body and ambient temperature, humidity, subject activity level, skin hydration, vasodilation, etc.). The implication is that, regardless of how reliable and accurate transdermal alcohol device hardware becomes at measuring TAC, the raw TAC data will never consistently map directly onto breath/blood alcohol concentration (BrAC/BAC) across individuals and drinking episodes. As a result, these devices have not seen wide-spread acceptance by the alcohol research and clinical communities where BAC/BrAC have been the standard measures of alcohol intoxication.

In our work to date on developing a means of estimating BrAC/BAC from TAC (see, for example, Dai, Rosen, Wang, Barnett, and Luczak (2016) and Rosen, Luczak, and Weiss (2014)), we have taken a deterministic approach to converting TAC to BAC or BrAC. First principles physics-based models (a one-dimensional diffusion equation with either infinite or finite speed of propagation

☆ This research was supported by grants from the National Institute on Alcohol Abuse and Alcoholism (NIAAA) (R21AA17711, R01AA026368-01 S.E.L./I.G.R.), and (R01AA025969, C.E.F.). The material in this paper was partially presented at the SIAM Conference on Uncertainty Quantification, April 16–19, 2018, Garden Grove, California, USA. This paper was recommended for publication in revised form by Associate Editor Martin Enqvist under the direction of Editor Torsten Söderström.

* Corresponding author.
*E-mail addresses:* sirlanci@caltech.edu (M. Sirlanci), luczak@usc.edu (S.E. Luczak), cfairbai@illinois.edu (C.E. Fairbairn), dkang38@illinois.edu (D. Kang), ruoxipan@usc.edu (R. Pan), yuxin@usc.edu (X. Yu), grosen@usc.edu (I.G. Rosen).

https://doi.org/10.1016/j.automatica.2019.04.026
0005-1098/

**Fig. 1.** Left: The WRISTAS™7 transdermal alcohol biosensor. Right: The Alcohol Monitoring System (AMS) Secure Continuous Alcohol Monitoring (SCRAM) system.

and input and output on the boundary) were fit to *individual calibration data* to capture the forward process — the transport of ethanol molecules from the blood, through the skin, and its measurement by the sensor. The result is TAC expressed as a convolution of BAC or BrAC with a kernel or filter. We then deconvolve an estimate of the BAC or BrAC from the biosensor measurements of TAC.

To reduce the burden on researchers/clinicians and participants/patients and thus significantly increase the feasibility of using these devices, we have been investigating ways to eliminate the need to calibrate the sensor's data analysis system to each individual, each sensor, and across varying current environmental conditions. In particular, we have been investigating the use of our first principles physics/physiological based models to describe the dynamics common to the entire population (*interpreted broadly to include all individuals, devices, and environmental conditions*) and then to attribute all un-modeled sources of uncertainty (primarily due to variations in physiology, hardware, and the environment) observed in individual data to random effects. Thus our *population model* is our deterministic transport equation with the parameters now being random. We estimate their distribution based on population data.

We assume an underlying mathematical framework describing the system dynamics that are common to all individuals, environmental conditions, and devices in the population (e.g., the physics-based model for the transport of ethanol from the blood, through the skin, and measurement by the sensor), but that individual members of the population exhibit variation in the parameters appearing in the model (e.g., the rate at which the alcohol is transported, evaporates, etc.). We then assume that the sensor measures the sum or mean of all of these effects. This is realized in the form of a model based on random partial differential equations and boundary conditions, and then, instead of fitting the unknown parameters in the model directly to individual data, we now estimate the distribution of the random parameters (in the form of probability measures or density functions) based on aggregate population data.

Consider the set of discrete time input–output systems

$$x_{j+1,i} = g(t_j, x_{j,i}, u_{j,i}; q_i), \quad x_{0,i} \text{ given} \tag{1.1}$$

$$y_{j,i} = h(t_j, x_{j,i}, u_{j,i}; q_i), \tag{1.2}$$

in an in general infinite dimensional Hilbert space $\mathscr{H}$, where $g : \mathbb{R}^+ \times \mathscr{H} \times \mathbb{R}^m \times Q \to \mathscr{H}$, $h : \mathbb{R}^+ \times \mathscr{H} \times R^m \times Q \to \mathbb{R}^\ell$, for $j = 0, \ldots, \mu_i$ $t_j = j\tau$, with $\tau > 0$ the length of the sampling interval, and for $i = 1, 2, \ldots, \nu$, the state and input for the $i$th system given by $x_i(q_i) = \{x_{j,i}(x_{0,i}, u_i, q_i)\}_{j=0}^{\mu_i}$ and $u_i = \{u_{j,i}\}_{j=0}^{\mu_i - 1}$, respectively.

In Eqs. (1.1), (1.2), it is assumed that the parameters $q_i$ are contained in a set $Q$, the set of admissible parameters, and that their values are specific to a particular individual (interpreted broadly) member of the population, individual $i$. The objective is to estimate the distribution of $q_i$'s, based on population or aggregate data rather than data for a particular individual. We assume $\mathscr{q}$ is a random vector with support set $Q$ and that $\mathscr{q} \sim \pi(\rho_0)$, for some $\rho_0 \in \mathscr{R}$ where $\pi(\rho)$ is a family of distributions or push forward measures parameterized by $\rho$ and defined on a sigma field of subsets of $Q$ where $\rho \in \mathscr{R}$, $\mathscr{R}$ the set of feasible parameters. We will assume $\pi(\rho)$ has a joint density function $f(\cdot; \rho)$; $\pi(A; \rho) = P(\mathscr{q} \in A) = \int_Q \chi_A d\pi(\rho) = \int_A f(q; \rho) dq$, for $A$ an event, where $\rho \in \mathscr{R}$. A comprehensive treatment of uncertainty quantification and statistical methodology in the context of inverse problems involving both finite and infinite dimensional dynamical systems that also includes an extensive survey and bibliography of the associated literature can be found in Banks, Hu, and Thompson (2014).

One way to formulate the underlying statistical model for estimating the parameters $\rho$ is to treat (1.1), (1.2) as an, in general, nonlinear mixed effects model (see, for example, Davidian and Giltinan (1995, 2003) and Demidenko (2013)) and find the maximum likelihood estimator for $\rho_0$. We assume that the observed data points are of the form $V_{j,i} = y_{j,i} + \varepsilon_{j,i}$, $j = 0, \ldots, \mu_i$, $i = 1, \ldots, \nu$, where $\varepsilon_{j,i}$, $j = 0, \ldots, \mu_i$, $i = 1, \ldots, \nu$, represent the measurement noise that are assumed to be independent across subjects (i.e. with respect to $i$), conditionally independent with respect to $\mathscr{q}$ within subjects (i.e. with respect to $j$), identically distributed with mean zero and variance $\sigma_i^2$, and with probability density function assumed to be continuous; to wit $\varepsilon_{j,i} \sim \varphi_i$, $j = 0, \ldots, \mu_i$, $i = 1, \ldots, \nu$. We note that it is possible to assume that the $\varphi_i$'s also depend on parameters that require estimation along with $\rho_0$ and the $\sigma_i^2$'s, but for simplicity we assume here that the densities $\varphi_i$ and variances $\sigma_i^2$ are known a-priori. Using conditional probability and the total probability formula, a likelihood function can be defined as $\mathscr{L}(\rho; V, v) = J_L(\rho; V, v) = \prod_{i=1}^{\nu} \int_Q L_i(q; V, v) f(q; \rho) dq = \prod_{i=1}^{\nu} \int_Q \prod_{j=0}^{\mu_i} \varphi_i(V_{j,i} - v_{j,i}) f(q; \rho) dq$, where $V = \{V_{j,i}\}$ and $v = v(q, u) = \{v_{j,i}\}$ with $v_{j,i} = h(t_j, x_{j,i}, u_{j,i}, q)$ and $u = \{u_{j,i}\}$. Once one deals with several computational issues including the finite dimensional approximation of the, in general, infinite dimensional state equation (1.1), the efficient evaluation of a potentially high dimensional integral, the loss of precision and underflow issues due to the fact that the evaluation of $\mathscr{L}$ requires the computation of potentially high order products of very small numbers, one could then seek a maximum likelihood estimator (MLE) for $\rho_0$ by maximizing $J_L$ over $\mathscr{R}$. Under appropriate regularity assumptions on the $\varphi_i$'s, $f$, and the system (1.1), (1.2), one way to do this might be to locate local maxima via gradient search. An alternative approach would be to seek a global max via some form of stochastic optimization. One could also treat direct observations of $\mathscr{q}$ as missing data and then use the iterative E-M algorithm to find the MLE (see, for example, Casella and Berger (2002)).

The likelihood function defined above could also serve as the basis for a Bayesian approach (see, for example, Calvetti, Kaipio, and Somersalo (2014), Dashti and Stuart (2017), Sirlanci, Rosen, Luczak, Fairbairn, Bresin, and Kang (2018) and Stuart (2010)). There are a couple of ways this can be done (see Sirlanci, Luczak, and Rosen (2019) and Sirlanci et al. (2018)). For example, assume a prior density $\mathscr{p}$ for $\rho$ and then apply Bayes to obtain the posterior, $\hat{\mathscr{p}}$, as $\hat{\mathscr{p}}(\rho) = \hat{\mathscr{p}}(\rho|V) = \frac{1}{Z}\mathscr{L}(\rho; V, v)\mathscr{p}(\rho)$, where $Z$ is given by $Z = \int_{\mathscr{R}} \mathscr{L}(\rho; V, v)\mathscr{p}(\rho) d\rho$. The Bayesian approach has some of the same computational challenges as the MLE when importance sampling is used to evaluate the integrals in the expression for $Z$ and an MCMC technique such as Metropolis–Hastings or the Gibbs Sampler is used to sample the posterior.

From a pharmacokinetic (PK) standpoint, in the alcohol biosensor problem, our definition of "subject" or population is somewhat non-standard in that it refers to not only individual participants and their various physiological differences, but also environmental conditions such as temperature and humidity, etc.

as well as hardware related uncertainties. Thus, in our model, the $q_i$'s describe unmodeled phenomena present both across different individuals and within the same individual. Indeed our working hypothesis is that each observation represents a sum or average of any number of diffusion processes all at work simultaneously. In addition, because human subjects are involved, it would be both costly and time consuming to collect simultaneous BrAC and TAC data from enough individuals wearing enough sensors and under sufficiently varied environmental conditions to estimate the distribution of the $q_i$'s directly. Although it can have bias problems (Mould & Upton, 2012), this suggests a third statistical model known in the PK literature as the *Naïve Pooled Data Estimator* (NPDE). For $\rho \in \mathscr{R}$ we define $v = v(\rho, u) = \{v_{j,i}\}$ by $v_{j,i} = \mathbb{E}_{\pi(\rho)}[y_{j,i}] = \int_Q h(t_j, x_{j,i}, u_{j,i}; q)d\pi(\rho) = \int_Q h(t_j, x_{j,i}, u_{j,i}; q)f(q;\rho)dq$, for $j = 0, 1, \ldots, \mu_i$, $i = 1, \ldots, \nu$. Then as is the case with standard regression, we now assume our observations are given by $V_{j,i} = \mathbb{E}_{\pi(\rho_0)}[y_{j,i}] + \varepsilon_{j,i}$, for $j = 0, 1, \ldots, \mu_i$, $i = 1, \ldots, \nu$, where the $\varepsilon_{i,j}$ represent measurement noise which are assumed to be independent, identically-distributed random variables with mean 0 and common variance $\sigma^2$, and $\rho_0 \in \mathscr{R}$ represents the *true* values of the parameters.

Our estimation problem minimizes prediction error in a nonlinear least squares sense. For $V = \{V_{j,i}\}$, the NPDE $\hat{\rho}$ in its most general form is given by

$$\hat{\rho} = \arg\min_{\rho \in \mathscr{R}} J_P(\rho; V, v), \tag{1.3}$$

with $J_P(\rho; V, v) = \sum_{i=1}^{\nu} J_{i,\mu}(\rho; V, v) + \lambda(\rho) = \sum_{i=1}^{\nu}\sum_{j=0}^{\mu_i} \alpha_{j,i}|v_{j,i} - V_{j,i}|^2 + \lambda(\rho)$, where $\lambda(\rho)$ is an optional regularization term, $\mu = \{\mu_i\}_{i=1}^{\nu}$, the $\alpha_{j,i}$'s are optional positive weights, and $v = \{v_{j,i}\} = \{v_{j,i}(\rho)\}$ and $V = \{V_{j,i}\}$ are as they were defined in the previous paragraph. We note that there is a relationship between the Bayesian estimator and the NPDE. If the $\{\varepsilon_{j,i}\}$ and the prior $\wp$ are assumed to be normal, the regularization $\lambda(\rho)$ in (1.3) is taken to be quadratic in $\rho$, and $\alpha_{j,i} = 1/2\sigma_i^2$, $j = 0, 1, \ldots, \mu_i$, $i = 1, \ldots, \nu$, then the NPDE, $\hat{\rho}$, is the so-called Maximum A-Posteriori or MAP estimator. Estimation problems involving Infinite dimensional or distributed parameter systems are often ill-posed in the sense that the unknown parameters either may not be, or may be close to not being, identifiable (see, for example, the treatment in Banks and Kunisch (2012), and the references contained therein). Mathematically, identifiability refers to the question of whether or not the map from the parameters to the data is injective, and if it is in fact invertible, whether or not the inverse map is continuous. The absence or near absence of identifiability can pose significant challenges when one attempts to solve the parameter estimation problem. Introducing regularization into the cost functional, along with compactness of the feasible parameter set, are ways of enhancing the identifiability of the parameters to be estimated. For example, Tychonoff regularization in the form of a regularization parameter or weight multiplied by the square of the norm of (a translate of) the parameters tends to increase the local convexity of the performance index thus making the existence of a unique minimizer more likely and easier to locate (Banks & Kunisch, 2012). The regularization term in (1.3) also often comes into play when one is attempting to estimate the *shape* of a functional parameter, which in our case here would be the unknown density $f$. In this case the parameters $\rho$ may be the coefficients in either (1) an expansion of the density $f$ in terms of a family of basis elements (e.g. piecewise polynomial splines) or (2) a polynomial chaos expansion of the random variable $q$ itself, for example, in terms of Hermite polynomials in the standard normal. In this case the feasible parameter set, $\mathscr{R}$, can be high dimensional and if the inverse of the parameter to data map is not continuous or unstable, over fitting (i.e. the fitting of noise in the data as opposed to the signal itself) can become an issue. In our treatment here we focus on the case where we are fitting parameterized exponential families (see, for example, Casella and Berger (2002)) of densities where the parameters $\rho$ are relatively low dimensional, for example the support, the mean vector, and the covariance matrix in a multi-variate normal (dimension on the order of ten), or the parameters $\alpha$ and $\beta$ in a gamma distribution. In the examples we looked at, the resulting optimization problems appeared to be well posed and we did not require regularization. However, once we estimated the distribution of the random parameters and used the resulting population model to estimate the BrAC input to the system, as a function of both time and the parameters, we found that regularization was absolutely essential and indispensable (see Sirlanci et al. (2018)).

The central focus of our work here is the finite dimensional approximation of the infinite dimensional state equation as given in (1.1) and the convergence of the corresponding estimated distribution for the random parameters when the necessary computations are based on this finite dimensional approximation. Moreover, we are only interested here in how these ideas apply to the alcohol biosensor problem described earlier. The underlying model in this case is based on diffusion with input and output on the boundary of the domain. Consequently we consider dynamical systems governed by abstract parabolic or regularly dissipative operators with unbounded input and output operators. Also, in the present effort, we focus only on the MLE and naïve pooled data estimator since they both involve the solution of an optimization problem. The Bayesian approach which is based on sampling will be treated elsewhere.

The abstract functional analytic approximation and convergence theory on which the results presented here are based was established in Sirlanci et al. (2019). In addition, in our treatment here, we are only concerned with the fitting of the random parameters in the forward model. The inverse problem involving the deconvolution of the BAC/BrAC from the TAC signal once the forward model with random elements has been fit is treated elsewhere (Sirlanci et al., 2018). Our general approach relies to some extent on two relatively recent papers: (1) Banks and Thompson's framework for the estimation of probability measures in random abstract evolution equations and the convergence of finite dimensional approximations in the Prohorov metric (Banks, Flores, Rosen, Rutter, Sirlanci, & Thompson, 2018; Banks & Thompson, 2012), and (2) Gittelson, Andreev, and Schwab's theory for random abstract parabolic partial differential equations with dynamics defined in terms of coercive sesquilinear forms (Gittelson, Andreev, & Schwab, 2014). The approach in Gittelson et al. (2014) is novel in the way that it treats the random variable as another "space-like" independent variable in the PDE. In this way, finite dimensional approximation is handled in much the same way that it is for the standard deterministic space variables. We use the framework in Gittelson et al. (2014) together with the generation and approximation results from linear semigroup theory, (i.e. the Hille–Yosida–Phillips and Trotter–Kato theorems, see, for example, Banks and Kunisch (2012), Kato (2013) and Pazy (1983)) to establish that the sufficient conditions for a Banks–Thompson-like convergence result are satisfied in the case of regularly dissipative systems with random parameters whose distributions can be described by appropriately parameterized probability density functions.

An outline of the paper is as follows. In Section 2 we discuss our diffusion-based model for the transdermal transport of ethanol, its abstract formulation as an infinite dimensional dynamical system with unbounded input and output, and its sampled form as a discrete-time input–output system. In Section 3 we discuss initial value problems involving regularly dissipative operators with random coefficients as in Gittelson et al. (2014), Schwab and Gittelson (2011), and apply it to the model derived in Section 2. Our finite-dimensional approximation and convergence results are in Section 4, Section 5 contains a consistency result for the estimator. In Section 6 we present numerical results using experimental/clinical drinking data, and Section 7 has a brief discussion of future research, and concluding remarks.

We use standard notation throughout. For example, we denote the space of square Lebesgue or Bochner integrable functions defined on an interval $(a, b)$ with range in the normed linear space $X$ by $L_2(a, b; X)$, and we use $C(a, b; X)$ when the functions are continuous. When $X = \mathbb{R}$, the range space is omitted. For normed linear spaces $X$ and $Y$, $\mathscr{L}(X, Y)$ denotes the space of bounded (continuous) linear operators defined on $X$ with range in $Y$. Unless explicitly stated otherwise, all Hilbert space norms are the ones induced by the standard inner product on that space. We occasionally use "dot" notation to denote weak or strong derivatives with respect to time.

## 2. A distributed parameter model for a transdermal alcohol biosensor and its abstract formulation

Our ethanol transport model is Fick's law-based. Consequently it is described by an abstract parabolic operator. In a Gelfand triple setting, these operators are examples of regularly dissipative operators that generate holomorphic semigroups (see Kato (2013), Pazy (1983) and Tanabe (1979)).

After converting to what are essentially dimensionless quantities (see Rosen et al. (2014)), we obtain the input/output model

$$\varphi_t(t, \eta) = q_1 \varphi_{\eta\eta}(t, \eta),\ 0 < \eta < 1,\ t > 0, \tag{2.1}$$

$$q_1 \varphi_\eta(t, 0) - \varphi(t, 0) = 0,\ t > 0, \tag{2.2}$$

$$q_1 \varphi_\eta(t, 1) = q_2 u(t),\ t > 0, \tag{2.3}$$

$$\varphi(0, \eta) = \varphi_0,\ 0 < \eta < 1, \tag{2.4}$$

$$y(t) = \varphi(t, 0),\ t > 0, \tag{2.5}$$

in the form of an initial–boundary value problem for a one dimensional diffusion equation with input and output on the boundary and two unknown parameters, $q = (q_1, q_2)$. In the system (2.1)–(2.5), $\varphi(t, \eta)$ is essentially the concentration of ethanol in the interstitial fluid in the epidermal layer of the skin at depth $\eta$ and time $t$, $u$ is the concentration of alcohol in the blood (BAC) as measured by a breath analyzer (BrAC), and $y$ is the (TAC). The boundary condition (2.2) models the evaporation of ethanol at the skin surface, condition (2.3) captures the exchange of ethanol molecules between the (blood-fed) dermal and epidermal layers of the skin. The output equation (2.5) models the biosensor measured TAC at the skin surface. We assume that there is no alcohol in the skin initially, so in general $\varphi_0 = 0$ in (2.4).

We use linear semigroup theory to reformulate (2.1)–(2.5) as a discrete-time SISO system in an infinite dimensional Hilbert space. In (2.1)–(2.5) the input and output are on the boundary, consequently the resulting input and output operators are unbounded with respect to the standard state space, $L_2(0, 1)$. In the discrete or sampled time formulation these operators become bounded.

Let $V$ and $H$ be Hilbert spaces with the embeddings $V \hookrightarrow H \hookrightarrow V^*$ dense and continuous, where $V^*$ denotes the space of continuous linear functionals on $V$. Let $\langle\cdot,\cdot\rangle$ and $|\cdot|$ denote the $H$ inner product and norm, respectively, and let $\|\cdot\|$ denote the norm on $V$. For $q \in \{Q, d_Q\}$, a compact metric space, let $a(q; \cdot, \cdot) : V \times V \to \mathbb{R}$ be a bilinear form. We assume:

(i) (Boundedness) $|a(q; \psi_1, \psi_2)| \le \alpha_0 \|\psi_1\| \|\psi_2\|$, $\psi_1, \psi_2 \in V$, $q \in Q$,

(ii) (Coercivity) $a(q; \psi, \psi) + \lambda_0 |\psi|^2 \ge \mu_0 \|\psi\|^2$, $\psi \in V$, $q \in Q$,

(iii) (Measurability) For $\psi_1, \psi_2 \in V$, the map $q \mapsto a(q; \psi_1, \psi_2)$ is measurable with respect to all measures $\pi(\rho)$, $\rho \in \mathscr{R}$.

For $q \in \{Q, d_Q\}$ let $b(q), c(q) \in V^*$ and consider an input/output system in weak form as given by

$$\begin{aligned} \langle \dot\varphi, \psi \rangle_{V^*,V} + a(q; \varphi, \psi) &= \langle b(q), \psi \rangle_{V^*,V} u, \\ y &= \langle c(q), \varphi \rangle_{V^*,V}, \quad \psi \in V, \end{aligned} \tag{2.6}$$

where $\varphi(0) = \varphi_0 \in H$ and $\langle\cdot,\cdot\rangle_{V^*,V}$ denotes the natural extension of the $H$ inner product to the duality pairing between $V$ and $V^*$. If we set $W(0, T) = \{\psi : \psi \in L_2(0, T; V), \dot\psi \in L_2(0, T; V^*)\}$ and $u \in L_2(0, T)$ it can be shown (see, for example, Lions (1971)) that the system (2.6) admits a unique solution $\varphi \in W(0, T)$ that depends continuously on $u \in L_2(0, T)$. It follows that $W(0, T) \subseteq C(0, T; H)$ and that $y \in L_2(0, T)$.

For $q \in Q$, the $q$-dependent bilinear form on $V \times V$, $a(q; \cdot, \cdot) : V \times V \to \mathbb{R}$, defines a bounded linear operator $A(q) \in \mathscr{L}(V, V^*)$ by $\langle A(q)\psi_1, \psi_2 \rangle_{V^*,V} = -a(q; \psi_1, \psi_2)$, for $\psi_1, \psi_2 \in V$. Then, if we let $\mathscr{H}$ denote any of the spaces $V, H$ or $V^*$, we can consider the linear operator $A(q)$ to be the unbounded operator, $A(q) : D_q \subset \mathscr{H} \to \mathscr{H}$ where $D_q = V$ in the case $\mathscr{H} = V^*$, and $D_q = \{\psi \in V : A(q)\psi \in \mathscr{H}\}$ in the case $\mathscr{H} = H$ or $\mathscr{H} = V$. It can be shown (see, for example, Banks and Ito (1988), Banks and Kunisch (2012) and Tanabe (1979)) that $A(q)$ is a closed, densely defined unbounded linear operator on $\mathscr{H}$ and it is the infinitesimal generator of an analytic semigroup of bounded linear operators, $\{e^{A(q)t} : t \ge 0\}$ on $\mathscr{H}$.

For $q \in Q$, define the bounded linear operators $B(q) : \mathbb{R} \to V^*$ and $C(q) : V \to \mathbb{R}$ by $\langle B(q)u, \psi \rangle_{V^*,V} = \langle b(q), \psi \rangle_{V^*,V} u$ and $C(q)\psi = \langle c(q), \psi \rangle_{V^*,V}$, respectively, for $\psi \in V$ and $u \in \mathbb{R}$. The input/output system can now be written *formally* in the standard state space form as

$$\dot x(t) = A(q)x(t) + B(q)u(t),\ y(t) = C(q)x(t), \tag{2.7}$$

for $t > 0$, where $x(0) = \varphi_0 \in H$ and the state $x(t) = \varphi(t, \cdot)$. Using the fact that $\{e^{A(q)t} : t \ge 0\}$ is an analytic semigroup on the spaces $V, H$ and $V^*$, it follows that a so-called mild solution (Kato, 2013; Pazy, 1983) to the state equation in (2.7), $x$ is given by $x(t; q) = e^{A(q)t}x_0 + \int_0^t e^{A(q)(t-s)}B(q)u(s)ds$, $t \ge 0$, $x \in W(0, T)$.

Let $\tau > 0$ and $x_0 = \varphi_0 \in V$ be given, and consider zero order hold inputs $u(t) = u_i$, $t \in [i\tau, (i+1)\tau)$, $i = 0, 1, 2 \ldots$. Under the assumptions we have made thus far (see Sirlanci et al. (2019)), using the variations of constants formula it can be shown that the properties of analytic semigroups generated by regularly dissipative operators yield

$$x_{i+1} = \hat A(q)x_i + \hat B(q)u_i,\ y_i = \hat C(q)x_i, \tag{2.8}$$

for $i = 0, 1, 2, \ldots$, where $\hat A(q) = e^{A(q)\tau}$, $\hat B(q) = \int_0^\tau e^{A(q)s}B(q)ds \in \mathscr{L}(\mathbb{R}, V)$, $\hat C(q) = C(q) = c(q) \in \mathscr{L}(V, \mathbb{R}) = V^*$, $x_i = x(i\tau)$ and $y_i = y(i\tau)$, $i = 0, 1, 2, \ldots$. Boundedness of $\hat A(q)$ and $\hat B(q)$ in (2.8) follows from the fact that $\{e^{A(q)t} : t \ge 0\}$ is an analytic semigroup on $V, H$ and $V^*$ (Banks & Ito, 1988; Banks & Kunisch, 2012; Lions, 1971; Tanabe, 1979). Coercivity, Assumption (ii) (possibly requiring a change of variables), implies without loss of generality that $A(q)$ from either $H$ into $H$ or $V$ into $V^*$ is invertible with bounded inverse and therefore that $\hat B(q) = \int_0^\tau e^{A(q)s}B(q)ds = A(q)^{-1}e^{A(q)s}B(q)|_0^\tau = (\hat A(q) - I)A(q)^{-1}B(q) = (\hat A(q) - I)A(q)^{-1}b(q) \in \mathscr{L}(\mathbb{R}, V)$.

Let $Q$ be a compact subset of $\mathbb{R}^2$ endowed with the Euclidean metric, let $H = L_2(0, 1)$ together with the standard inner product $\langle \psi_1, \psi_2 \rangle = \int_0^1 \psi_1\psi_2 d\eta$, and norm denoted by $|\cdot|$, and let $V$ be the Sobolev space $H^1(0, 1)$ together with its standard inner product $\langle\langle \psi_1, \psi_2 \rangle\rangle = \int_0^1 \psi_1\psi_2 d\eta + \int_0^1 \psi_1'\psi_2' d\eta$ and norm denoted by $\|\cdot\|$. The usual dense and continuous embeddings $V \hookrightarrow H \hookrightarrow V^*$ hold, where $V^* = H^{-1}(0, 1)$ denotes the dual space of $V$. The forms and functionals $a(q;\cdot,\cdot) : V\times V \to \mathbb{R}$, $b(q;\cdot) : V \to \mathbb{R}$, and $c(q;\cdot) : V \to \mathbb{R}$ are given by

$$a(q;\psi_1,\psi_2) = \psi_1(0)\psi_2(0) + q_1\int_0^1 \psi_1'\psi_2' d\eta, \tag{2.9}$$

for $\psi_1, \psi_2 \in V$, $\langle b(q), \psi\rangle_{V^*,V} = q_2\psi(1)$, and $\langle c(q), \psi\rangle_{V^*,V} = \psi(0)$, for $\psi \in V$. It follows that $b(q) = q_2\delta(\cdot - 1) \in V^*$ and $c(q) = \delta \in V^*$, where $\delta$ denotes the Dirac delta distribution, or unit impulse at zero. It is not difficult to argue that Assumptions (i)–(iii) hold for the form $a(\cdot;\cdot,\cdot)$ as given in (2.9). See Sirlanci et al. (2019) for a more abstract, detailed and rigorous description of how to deal with input signals on the boundary of the domain.

## 3. Systems governed by regularly dissipative operators with random parameters

We use ideas from Gittelson et al. (2014) to consider systems of the form (2.1)–(2.5) with the parameters $q \in Q$ random. Let $\mathscr{q}$ be a $p$ dimensional random vector with support in $\prod_{i=1}^p [a_i, b_i]$ where $-\infty < \bar{\alpha} < a_i < b_i < \bar{\beta} < \infty$. Let $\vec{a}, \vec{b} \in \mathbb{R}^p$ be given by $\vec{a} = [a_i]$ and $\vec{b} = [b_i]$, and let $\Theta$ be a parameter set that is a compact subset of $\mathbb{R}^r$ for some $r$. We assume that $\mathscr{q}$ has distribution described by the absolutely continuous cdf, $F(q;[\vec{a},\vec{b},\theta])$, or equivalently by the push forward measure, $\pi = \pi([\vec{a},\vec{b},\theta])$, where $\theta \in \Theta$.

Let $a(q;\cdot,\cdot)$ denote a sesquilinear form on $V\times V$ satisfying the conditions (i)–(iii) given in Section 2, and in particular that the function $q \mapsto a(q; v, w)$ is $\pi$-measurable for any $v, w \in V$. By appropriately defining new function spaces, it is possible to embed the randomness in the sesquilinear form $a(\mathscr{q};\cdot,\cdot)$, or equivalently in the operator, $A(\mathscr{q})$ in (2.7) into these spaces. In effect, the random parameters are treated as if they were space variables in the underlying PDE. The input output system (2.7) now becomes amenable to analysis and approximation via linear semigroup theory.

We define $\mathscr{V} = L^2_\pi(Q;V)$ and $\mathscr{H} = L^2_\pi(Q;H)$. It then follows that $\mathscr{V}, \mathscr{H}, \mathscr{V}^*$ form a Gelfand triple of separable Hilbert spaces with $\mathscr{V} \subseteq \mathscr{H} \subseteq \mathscr{V}^*$ by identifying $\mathscr{H}$ with its dual $\mathscr{H}^*$ and identifying $\mathscr{V}^*$ with the Bochner space $L^2_\pi(Q;V^*)$. We also define the $\pi$-averaged sesquilinear form $\mathscr{a}(v,w) : \mathscr{V}\times\mathscr{V} \to \mathbb{C}$ by

$\mathscr{a}(v,w) = \int_Q a(q; v(q), w(q))d\pi = \mathbb{E}_\pi[a(\mathscr{q}; v(\mathscr{q}), w(\mathscr{q}))]$,

where, $v, w \in \mathscr{V}$. Assumptions (i) and (ii) guarantee that $\mathscr{a}(\cdot,\cdot)$ given above is a bounded and coercive sesquilinear form on $\mathscr{V}\times\mathscr{V}$, and therefore that $-\mathscr{a}$ induces a bounded linear map $\mathscr{A}$ from $\mathscr{V}$ into $\mathscr{V}^*$. It follows (see, once again, for example, Banks and Ito (1988), Banks and Kunisch (2012) and Tanabe (1979)) that $\mathscr{A} : Dom(\mathscr{A}) \subseteq \mathscr{H} \to \mathscr{H}$ with $Dom(\mathscr{A}) = \{v \in \mathscr{V} : \mathscr{A}v \in \mathscr{H}\}$ is the generator of an analytic semigroup of bounded linear operators $\mathscr{T} = \{\mathscr{T}(t) : t \geq 0\}$ on $\mathscr{H}$ and that $\mathscr{T}$ can be extended to an analytic semigroup on $\mathscr{V}^*$ and restricted to an analytic semigroup on $\mathscr{V}$.

We assume that the maps $q \mapsto \langle b(q), \psi(q)\rangle_{V^*,V}$ and $q \mapsto \langle c(q), \psi(q)\rangle_{V^*,V}$ are $\pi$-measurable for any $\psi \in \mathscr{V}$, and that $\|b(q)\|_{V^*}$ and $\|c(q)\|_{V^*}$ are uniformly bounded for $q \in Q$. (Assuming that $b, c \in \mathscr{V}^*$ would be fine as well.) Define the bounded linear operators $\mathscr{B} : \mathbb{R} \to \mathscr{V}^*$ and $\mathscr{C} : \mathscr{V} \to \mathbb{R}$ by $\langle \mathscr{B}u, v\rangle = \int_Q \langle b(q), v(q)\rangle_{V^*,V} d\pi(q)u = \mathbb{E}_\pi[\langle b(\mathscr{q}), v(\mathscr{q})\rangle_{V^*,V}]u$ and $\mathscr{C}v = \int_Q \langle c(q), v(q)\rangle_{V^*,V} d\pi(q) = \mathbb{E}_\pi[\langle c(\mathscr{q}), v(\mathscr{q})\rangle_{V^*,V}]$, respectively, for $v \in \mathscr{V}$, $u \in \mathbb{R}$. For $q \in Q$, set $x_0(q) = \varphi_0$, constant with respect to $q$, in either $\mathscr{H}$ or $\mathscr{V}$.

We then consider the continuous time input/output system set in the spaces $\mathscr{V}, \mathscr{H}, \mathscr{V}^*$, given by

$$\dot{x}(t) = \mathscr{A}x(t) + \mathscr{B}u(t), \quad x(0) = x_0 \in \mathscr{H} \tag{3.1}$$

$$y(t) = \mathscr{C}x(t), \quad t > 0. \tag{3.2}$$

The mild solution to the initial value problem (3.1)–(3.2) can then be obtained via the variation of parameters formula as $x(t) = \mathscr{T}(t)x_0 + \int_0^t \mathscr{T}(t-s)\mathscr{B}u(s)ds$, from which together with (3.2) it follows that for $t \geq 0$, $y(t) = \mathscr{C}\mathscr{T}(t)x_0 + \int_0^t \mathscr{C}\mathscr{T}(t-s)\mathscr{B}u(s)ds$.

With sampling time $\tau > 0$, zero order hold inputs $u(t) = u_j$, $t \in [j\tau, (j+1)\tau)$, $x_j = x(j\tau)$, and $y_j = y(j\tau)$, $j = 0, 1, 2, \ldots$, variation of parameters yields

$$x_{j+1} = \hat{\mathscr{A}}x_j + \hat{\mathscr{B}}u_j, \;\; y_j = \hat{\mathscr{C}}x_j, \quad j = 0, 1, 2, \ldots \tag{3.3}$$

with $x_0 \in \mathscr{V}$ where $\hat{\mathscr{A}} = \mathscr{T}(\tau) \in \mathscr{L}(\mathscr{V},\mathscr{V})$, $\hat{\mathscr{B}} = \int_0^\tau \mathscr{T}(s)\mathscr{B}ds \in \mathscr{L}(\mathbb{R},\mathscr{V})$, and $\hat{\mathscr{C}} = \mathscr{C} \in \mathscr{L}(\mathscr{V},\mathbb{R}) = \mathscr{V}^*$. Boundedness of the operators $\hat{\mathscr{A}}$ and $\hat{\mathscr{B}}$ follows from the fact that $\{\mathscr{T}(t) : t \geq 0\}$ is an analytic semigroup on $\mathscr{V}, \mathscr{H}$ and $\mathscr{V}^*$ (Banks & Ito, 1988; Banks & Kunisch, 2012; Lions, 1971; Tanabe, 1979). Once again, without loss of generality, we may assume that $\mathscr{A} : Dom(\mathscr{A}) \subset \mathscr{V}^* \to \mathscr{V}^*$ is invertible with bounded inverse, from which it follows that $\hat{\mathscr{B}} = \int_0^\tau \mathscr{T}(s)\mathscr{B}ds = \mathscr{A}^{-1}\mathscr{T}(s)\mathscr{B}|_0^\tau = (\hat{\mathscr{A}} - I)\mathscr{A}^{-1}\mathscr{B} \in \mathscr{L}(\mathbb{R},\mathscr{V})$. Since $\mathrm{Dom}(\hat{\mathscr{B}}) = \mathbb{R}$, $\hat{\mathscr{B}} = \hat{\mathscr{b}}$, for some $\hat{\mathscr{b}} \in \mathscr{V}$.

It can be shown (Gittelson et al., 2014; Schwab & Gittelson, 2011) that $x(t)$ given by (3.1) agrees $\pi$-almost all $q \in Q$ with $x(t)$ given in (2.7) for all $t \geq 0$, from which it then immediately follows that for $\pi$-almost all $q \in Q$ $x_j$ given in (3.3) agrees with $x_j$ given in (2.8), $j = 0, 1, 2, \ldots$. Moreover, it follows that $y(t) = \mathscr{C}x(t) = \mathbb{E}_\pi[y(t;\mathscr{q})] = \mathbb{E}_\pi[C(\mathscr{q})x(t;\mathscr{q})]$, for all $t \geq 0$ and therefore that for $j = 0, 1, 2, \ldots$,

$$y_j = \hat{\mathscr{C}}x_j = \mathbb{E}_\pi[y_i(\mathscr{q})] = \mathbb{E}_\pi[\hat{C}(\mathscr{q})x_j(\mathscr{q})]. \tag{3.4}$$

The *population model* for the NPDE statistical model is then (3.1), (3.2) or (3.3), (3.4). For the mixed effects statistical model and the MLE we define the alternate continuous time output operator $y(t) = \mathscr{C}x(t) = C(\cdot)x(t,\cdot)$ with its discrete time counterpart defined as it was earlier simply as $\hat{\mathscr{C}} = \mathscr{C}$. In this case it will follow that for $\pi$-almost all $q \in Q$, $y(t)$ given in (3.2) agrees with $y(t)$ given in (2.7) for all $t \geq 0$ and $y_j$ given in (3.3) agrees with $y_j$ given in (2.8), for $j = 0, 1, 2, \ldots$, for $\pi$-almost all $q \in Q$.

## 4. Finite-dimensional approximation, convergence and computational considerations

We formulate our estimation/optimization problem (1.3) as follows:

(**P**) Given the sampled data $(\tilde{u}_i, \tilde{y}_i)_{i=1}^\nu = (\{\tilde{u}_{j,i}\}_{j=0}^{\mu_i-1}, \{\tilde{y}_{j,i}\}_{j=0}^{\mu_i})_{i=1}^\nu$, determine $\rho^* = [\vec{a}^*, \vec{b}^*, \theta^*]$ with $\vec{a}^* = [a_i^*]$, $\vec{b}^* = [b_i^*]$ and $\theta^* \in \Theta$, in the set of feasible parameters $\mathscr{R}$, which minimizes (maximizes) $J_P(\rho; V, y(\rho))$ $(J_L(\rho; V, y(\rho)))$ where $y(\rho) = (\{y_{j,i}(\tilde{u}_i,\rho)\}_{j=0}^{\mu_i-1})_{i=1}^\nu = (\{y_{j,i}(\tilde{u}_i,[\vec{a},\vec{b},\theta])\}_{j=0}^{\mu_i-1})_{i=1}^\nu$ is given by the appropriate population model as in (3.3) with $u_i = \{u_{j,i}\}_{j=0}^{\mu_i-1} = \{\tilde{u}_{j,i}\}_{j=0}^{\mu_i-1}$, $i = 1, 2, \ldots, \nu$, and $V = (\{\tilde{y}_{j,i}\}_{j=0}^{\mu_i})_{i=1}^\nu$

Henceforth we assume that the measures $\pi = \pi(\rho) = \pi([\vec{a},\vec{b},\theta])$ are described by a family of joint density functions $f = f(\cdot;\rho) = f(\cdot;[\vec{a},\vec{b},\theta])$ and let $\bar{Q} = \prod_{i=1}^p [\bar{\alpha},\bar{\beta}]$. We require the following assumptions on the family of densities $f(\cdot;\rho) = f(\cdot;[\vec{a},\vec{b},\theta])$:

(iv) The maps $[\vec{a},\vec{b},\theta] \mapsto f(q;[\vec{a},\vec{b},\theta])$ are continuous on $\mathbb{R}^p \times \mathbb{R}^p \times \Theta$ for $\pi$-almost every $q \in \bar{Q} = \prod_{i=1}^p [\bar{\alpha},\bar{\beta}]$.

(v) There exist positive constants $\gamma, \delta$ such that $0 < \gamma \leq f(q;[\vec{a},\vec{b},\theta]) \leq \delta < 0$,for $\pi$-almost every $q \in \bar{Q} = \prod_{i=1}^p [\bar{\alpha},\bar{\beta}]$.

Our discussions to follow apply to both the optimization problems stated in Problem (**P**). Consequently in what follows, we simply use $J$ to denote either $J_P$ or $J_L$. Assumptions (i)–(v) are sufficient to establish that the maps $\rho = [\vec{a},\vec{b},\theta] \mapsto y_{j,i}(\tilde{u}_i,[\vec{a},\vec{b},\theta])$ are continuous for $j = 0, 1, 2, \ldots, \mu_i - 1$ and $i = 1, 2, \ldots, \nu$, and therefore that the map $\rho = [\vec{a}, \vec{b}, \theta] \mapsto J([\vec{a}, \vec{b}, \theta]; V, y(\rho)) = J(\rho, V, y(\rho))$ is continuous. Consequently it follows from compactness that problem (**P**) has a solution.

Solving problem (**P**) requires finite dimensional approximation. Let $\bar{\mathscr{H}} = L^2_\pi(\bar{Q}; H)$, and $\bar{\mathscr{V}} = L^2_\pi(\bar{Q}; V)$. For $N = 1, 2, \ldots$, let $\vec{a}^N = [a_i^N]$ and $\vec{b}^N = [b_i^N]$ be vectors in $\mathbb{R}^p$ with $-\infty < \bar{\alpha} < a_i^N < b_i^N < \bar{\beta} < \infty$, and let $\theta^N \in \Theta$, and set $Q^N = \prod_{i=1}^p [a_i^N, b_i^N]$, $\mathscr{H}^N = L^2_\pi(Q^N; H)$, and $\mathscr{V}^N = L^2_\pi(Q^N; V)$. Let $\mathscr{U}^N$ be a finite dimensional subspace of $\mathscr{V}^N$. Let $\mathscr{I}^N : \bar{\mathscr{H}} \to \mathscr{H}^N$ be such that $Im(\mathscr{I}^N) = \mathscr{H}^N$ and $|\mathscr{I}^N x|_{\mathscr{H}^N} \le |x|_{\bar{\mathscr{H}}}$. Let $\mathscr{P}^N : \mathscr{H}^N \to \mathscr{U}^N$ be the orthogonal projection of $\mathscr{H}^N$ onto $\mathscr{U}^N$ and define $\mathscr{J}^N := \mathscr{P}^N \circ \mathscr{I}^N$. Let $v^N, w^N \in \mathscr{U}^N$ and define $\mathscr{A}^N : \mathscr{U}^N \to \mathscr{U}^N$ by $\langle \mathscr{A}^N v^N, w^N \rangle = -a(v^n, w^N) = -\int_{Q^N} a(q; v^N(q), w^N(q)) d\pi^N(q) = -\int_{Q^N} a(q; v^N(q), w^N(q)) f(q; [\vec{a}^N, \vec{b}^N, \theta^N]) dq$.

Using assumptions (i)–(v), it can then be shown (see Sirlanci et al. (2019, 2018)) that $\mathscr{A}^N \in G(M, \lambda_0)$ on $\mathscr{U}^N$ (i.e., $|e^{\mathscr{A}^N t}|_{\mathscr{H}^N} \le M e^{\lambda_0 t}$, $t \ge 0$), where the constants $M$ and $\lambda_0$ are independent of $N$. Suppose further that for some $\lambda \ge \lambda_0$, $|\mathscr{J}^N R_\lambda(\mathscr{A}) x - R_\lambda(\mathscr{A}^N) \mathscr{J}^N x|_{\mathscr{H}^N} \to 0$ as $N \to \infty$, for every $x \in \bar{\mathscr{H}}$, where $R_\lambda(\mathscr{A})$ and $R_\lambda(\mathscr{A}^N)$ denote respectively the resolvent operators of $\mathscr{A}$ and $\mathscr{A}^N$ at $\lambda$. A version (Banks, Burns, & Cliff, 1981; Sirlanci et al., 2019, 2018) of the Trotter Kato theorem (see, for example, Kato (2013) and Pazy (1983)) that allows for the state spaces to depend on the parameters $[\vec{a}^N, \vec{b}^N, \theta^N]$, implies that $|\mathscr{J}^N \mathscr{T}(t) x - e^{\mathscr{A}^N t} \mathscr{J}^N x|_{\mathscr{H}^N} \to 0$ as $N \to \infty$, for $x \in \bar{\mathscr{H}}$, uniformly in $t$ on compact intervals of $[0, \infty)$.

Define $\mathscr{B}^N : \mathbb{R} \to \mathscr{U}^N$ and $\mathscr{C}^N : \mathscr{U}^N \to \mathbb{R}$ by $\langle \mathscr{B}^N u, v^N \rangle = \int_{Q^N} \langle b(q), v^N(q) \rangle_{V^*,V} f(q; [\vec{a}^N, \vec{b}^N, \theta^N]) dq u$ and $\mathscr{C}^N v^N = \int_{Q^N} \langle c(q), v^N(q) \rangle_{V^*,V} f(q; [\vec{a}^N, \vec{b}^N, \theta^N]) dq$, where $v^N \in \mathscr{U}^N$ and $u \in \mathbb{R}$ and consider the sequence of finite dimensional optimization problems:

($\mathbf{P_N}$) Given the sampled data $(\tilde{u}_i, \tilde{y}_i)_{i=1}^{\nu} = \left(\{\tilde{u}_{j,i}\}_{j=0}^{\mu_i - 1}, \{\tilde{y}_{j,i}\}_{j=0}^{\mu_i}\right)_{i=1}^{\nu}$, determine $\rho^{N*} = [\vec{a}^{N*}, \vec{b}^{N*}, \theta^{N*}]$ with $\vec{a}^{N*} = [a_i^{N*}]$, $\vec{b}^{N*} = [b_i^{N*}]$ and $\theta^{N*} \in \Theta$ feasible (i.e. $\rho^{N*} \in \mathscr{R}$), which minimizes (maximizes) $J_P^N(\rho; V, y(\rho)) = J_P(\rho; V, y^N(\rho))$ ($J_L^N(\rho; V, y(\rho)) = J_L(\rho; V, y^N(\rho))$) where $y^N(\rho) = (\{y^N_{j,i}(\tilde{u}_i, [\vec{a}, \vec{b}, \theta])\}_{j=0}^{\mu_i})_{i=1}^{\nu}$ is given by

$$x^N_{j+1,i} = \hat{\mathscr{A}}^N x^N_{j,i} + \hat{\mathscr{B}}^N \tilde{u}_{j,i}, \; y^N_{j,i} = \mathscr{C}^N x^N_{j,i}, \tag{4.1}$$

with $x^N_{0,i} \in \mathscr{U}^N$ where $\hat{\mathscr{A}}^N = e^{\mathscr{A}^N \tau} \in \mathscr{L}(\mathscr{U}^N, \mathscr{U}^N)$, $\hat{\mathscr{B}}^N = \int_0^\tau e^{\mathscr{A}^N s} \mathscr{B}^N ds \in \mathscr{L}(\mathbb{R}, \mathscr{U}^N)$, $\hat{\mathscr{C}}^N = \mathscr{C}^N$ in $\mathscr{L}(\mathscr{U}^N, \mathbb{R})$ ($\mathscr{L}(\mathscr{U}^N, L_2(Q))$), and $V = \{\{\tilde{y}_{j,i}\}_{j=0}^{\mu_i}\}_{i=1}^{\nu}$

We have the following convergence theorem (see Sirlanci et al. (2019)).

**Theorem 4.1.** *Let assumptions (i)–(v) hold. Then for each $N = 1, 2, \ldots$, the problems ($\mathbf{P_N}$) given above admit a solution $\rho^{N*} = (\vec{a}^{N*}, \vec{b}^{N*}, \theta^{N*})$. Suppose further that*

*(vi) $|\mathscr{P}^N v^N - v^N|_{\mathscr{V}^N} \to 0$ as $N \to \infty$, $v^N \in \mathscr{V}^N$.*

*Then there exists a subsequence, $\{\rho^{N_j*}\}_{j=1}^\infty$, with $\rho^{N_j*} = (\vec{a}^{N_j*}, \vec{b}^{N_j*}, \theta^{N_j*}) \to (\vec{a}^*, \vec{b}^*, \theta^*) = \rho^*$ as $j \to \infty$, and $\rho^* = (\vec{a}^*, \vec{b}^*, \theta^*)$ is a solution to problem* (**P**).

The problems ($\mathbf{P_N}$) are solved numerically using an iterative gradient-based scheme. Once a basis for $\mathscr{U}^N$ is chosen, the operators in (4.1) can be represented as matrices and the cost functional $J = J_P$ or $J = J_L$ and its gradient can be computed. If for $N = 1, 2, \ldots$, $\mathscr{U}^N = span\{\psi_j^N\}_{j=1}^{K^N} \subset \mathscr{V}^N$, then the matrix representation for $\mathscr{A}^N \in \mathscr{L}(\mathscr{U}^N, \mathscr{U}^N)$ is given by $[\mathscr{A}^N]_{i,j} = -[\langle \psi_i^N, \psi_j^N \rangle_{\mathscr{H}^N}]^{-1} [a(\psi_i^N, \psi_j^N)]$, for $j = 1, 2, \ldots, K^N$. Matrix representations for the operators $\mathscr{B}^N$ and $\mathscr{C}^N$ are computed analogously. The matrix representations for $\mathscr{A}^N$, $\mathscr{B}^N$, and $\mathscr{C}^N$ can then be used in a straight forward manner to compute the matrix representations for the operators $\hat{\mathscr{A}}^N$, $\hat{\mathscr{B}}^N$, and $\hat{\mathscr{C}}^N$ appearing in (4.1). In each iteration of the gradient search, $\tilde{\nabla} J^N$ can be computed efficiently and exactly (i.e. without any truncation error) using the adjoint method with the requisite derivatives of the operators $\hat{\mathscr{A}}^N$, $\hat{\mathscr{B}}^N$, and $\hat{\mathscr{C}}^N$ computed efficiently using the sensitivity equations corresponding to (4.1) (see, for example, Levi and Rosen (2010) and Sirlanci et al. (2019)).

## 5. Consistency of the estimator

The NPDE $\hat{\rho}$ in (1.3) is given by $\hat{\rho} = [\vec{a}^*, \vec{b}^*, \theta^*]$, where $\vec{a}^* = [a_i^*]$, $\vec{b}^* = [b_i^*]$, and $\theta^* \in \Theta$ are a solution to problem (**P**) with $J = J_P$. Assumptions (a)–(e) below and Theorem 4.2 in Banks and Thompson (2012) allow us to establish the consistency of $\hat{\rho}$. An analogous consistency result for the MLE can also be proven, but the argument is not based on Theorem 4.2 in Banks and Thompson (2012). Consequently it will be discussed elsewhere.

(a) The measurement noise $\{\varepsilon_{i,j}\}$ is i.i.d. with respect to a probability space $\{\Omega, \Sigma, P\}$ with $\mathbb{E}_P[\varepsilon_{i,j}] = 0$ and $Var_P(\varepsilon_{i,j}) = \sigma^2$.
(b) The feasible set $\mathscr{R}$ is compact with $\mathrm{int}(\mathscr{R}) \neq \emptyset$.
(c) For $i = 1, 2, \ldots, \nu$, $\mu_i = \mu$, and $\mu\tau = T$, for some positive integer $\mu$ and some $T > 0$, where $\tau$ is the sampling time defined in Section 3.
(d) In problem (**P**) and the statistical model for the NPDE, we have $V_{j,i} = \tilde{y}_{i,j} = \mathbb{E}_{\pi(\rho_0)}[y_{j,i}] + \varepsilon_{j,i}$, $j = 0, 1, \ldots, \mu_i$, $i = 1, \ldots, \nu$ for some $\rho_0 \in int(\mathscr{R})$.
(e) If for each $i = 1, 2, \ldots, \nu$, $\tilde{y}(\cdot; \rho)$ is given by (3.2) with $u = \tilde{u}_i$ and $\pi = \pi(\rho)$, then $\rho_0$ is the unique minimizer of $J_{i,0}$ in $\mathscr{R}$, where for $\rho \in \mathscr{R}$, $J_{i,0}(\rho) = \sigma^2 + \int_0^t (\tilde{y}(t; \rho_0) - \tilde{y}(t; \rho))^2 dt$.

Theorem 4.2 in Banks and Thompson (2012) yields the following lemma.

**Lemma 5.1.** *Assume (i)–(v) and (a)–(e) hold. Then there is an $A \in \Sigma$ with $P(A) = 1$ such that for all $\omega \in A$ and $J_{i,\mu}(\rho; V, v)$ as given in the definition of $J_P$ following (1.3) with $\alpha_{j,i} = 1$ and $\lambda(\rho) = 0$, we have $\frac{1}{\nu}\sum_{i=1}^{\nu}\left\{\frac{1}{\mu} J_{i,\mu}(\rho; V(\omega), y(\rho)) - J_{i,0}(\rho)\right\} \to 0$, as $\nu, \mu \to \infty$, $\tau \to 0$ with $\mu\tau = T$, uniformly in $\rho$, for $\rho \in \mathscr{R}$.*

**Theorem 5.1** (*Consistency of the NPDE, $\hat{\rho}$*). *Let $\hat{\rho} \in \mathscr{R}$ be as in (1.3) and Problem* (**P**)*. Then under the assumptions of Lemma 5.1, $\hat{\rho} = [\vec{a}^*, \vec{b}^*, \theta^*]$ is consistent for $\rho_0$. That is $\hat{\rho} \xrightarrow{P} \rho_0$, as $\nu, \mu \to \infty$, $\tau \to 0$, with $\mu\tau = T$.*

**Proof.** As in the proof of Theorem 4.3 in Banks and Thompson (2012), for $\nu = 1, 2, \ldots$, let $J_0(\rho) = \frac{1}{\nu}\sum_{i=1}^{\nu} J_{i,0}(\rho)$ and $A \in \Sigma$ be as in the statement of Lemma 5.1. Let $\omega \in A$ be fixed, and let $\delta > 0$ be arbitrary. Then Lemma 5.1 implies that there exists an open local neighborhood of $\rho_0$ of radius $\delta$, $N_\delta(\rho_0)$, such that by Assumption (e) there exists an $\epsilon > 0$ for which $J_0(\rho) - J_0(\rho_0) > \epsilon$, for all $\rho \in \mathscr{R}_\delta$, where $\mathscr{R}_\delta$ is the compact (i.e., closed and bounded) set given by $\mathscr{R} \cap N_\delta(\rho_0)^c$. Now once again by Lemma 5.1, there exist $\nu_0$, $\mu_0$ and $\tau_0$, such that for all $\nu > \nu_0$, $\mu > \mu_0$ and $\tau < \tau_0$ with $\mu\tau = T$, $\left|\frac{1}{\nu\mu} J_P(\rho; V(\omega), y(\rho)) - J_0(\rho)\right| < \frac{\epsilon}{4}$, $\rho \in \mathscr{R}$, where $J_P$ is as given in (1.3) and Problem (**P**). Then with $\nu > \nu_0$, $\mu > \mu_0$, $\tau < \tau_0$, $\mu\tau = T$ and $\rho \in \mathscr{R}_\delta$, we have $\frac{1}{\nu\mu}(J_P(\rho; V(\omega), y(\rho)) - J_P(\rho_0; V(\omega), y(\rho_0))) = \frac{1}{\nu\mu} J_P(\rho, V(\omega), y(\rho)) \mp J_0(\rho) \mp J_0(\rho_0) - \frac{1}{\nu\mu} J_P(\rho_0, V(\omega), y(\rho_0)) \ge -\frac{\epsilon}{4} + \epsilon - \frac{\epsilon}{4} > 0$

But $J_P(\hat{\rho}, V(\omega), y(\hat{\rho})) \le J_P(\rho_0, V(\omega), y(\rho_0))$. It follows that $\hat{\rho} \in N_\delta(\rho_0)$ if $\nu > \nu_0$, $\mu > \mu_0$, $\tau < \tau_0$ with $\mu\tau = T$. Since $\delta > 0$ was arbitrary and $P(A) = 1$, we have that $\hat{\rho} \xrightarrow{P} \rho_0$, as $\nu, \mu \to \infty$, $\tau \to 0$ with $\mu\tau = T$. □

## 6. Numerical results

We present some numerical studies for the NPDE. Numerical results for the MLE are in progress and will be reported on elsewhere. The finite dimensional subspaces $\mathscr{U}^N$ were constructed as follows based on discretization of $(\eta, q_1, q_2)$-space. For each $n = 1, 2, \ldots$, let $\{\varphi_j^n\}_{j=0}^n$ denote the set of standard linear B-splines on the interval $[0, 1]$ defined with respect to the usual uniform mesh, $\{j/n\}_{j=0}^n$, and set $H^n = span\{\varphi_j^n\}_{j=0}^n \subset V$ (note the $\varphi_j^n$ are the usual "pup tent" or "chapeau" functions of height one and support of width $2/n$, $\left[\frac{j-1}{n}, \frac{j+1}{n}\right] \cap [0, 1]$). If $P^n : H \to H^n$ denotes the orthogonal projection of $H = L_2(0, 1)$ onto $H^n$, it is well known (see for example, Schultz (1972) that $\lim_{n\to\infty} P^n\varphi = \varphi$ in $H$ for $\varphi \in H$ and in $V$ for $\varphi \in V$. Then for $i = 1, 2$, and each $m_i = 1, 2, \ldots$ let $\left\{\chi_{i,j}^{m_i}\right\}_{j=1}^{m_i}$ denote the set of standard 0th order B-splines (i.e., piecewise constant functions) on the interval $[a_i, b_i]$ defined with respect to the usual uniform mesh, $\{a_i + (b_i - a_i)j/m_i\}_{j=0}^{m_i}$. If $P_i^{m_i}$ denotes the orthogonal projection of $L_2(a_i, b_i)$, it is not difficult to show that $\lim_{m_i\to\infty} P_i^{m_i}\zeta = \zeta$ in $L_2(a_i, b_i)$ for every $\zeta \in L_2(a_i, b_i)$. Then let $N$ denote the triple $(n, m_1, m_2)$, and $L$ the multi-index $(j, j_1, j_2)$, where $j \in \{0, 1, 2, \ldots, n\}$, $j_1 \in \{1, 2, \ldots, m_1\}$, and $j_2 \in \{1, 2, \ldots, m_2\}$. We use tensor products to define $\{\psi_L^N\}_L^N$ as $\psi_L^N = \varphi_j^n \chi_{1,j_1}^{m_1} \chi_{2,j_2}^{m_2}$ and set $\mathscr{U}^N = span\{\psi_L^N\}_L^N$. It can be argued that (vi) holds.

In what follows we fit a truncated bivariate normal. Let $\vec{q} = [q_1, q_2]$, $\vec{a} = [a_1, a_2]$, $\vec{b} = [b_1, b_2]$, and let $\phi(\cdot; \vec{\mu}, \Sigma)$ denote the joint density for the bivariate normal with mean $\vec{\mu}$ and covariance matrix $\Sigma$: $\phi(\vec{q}; \vec{\mu}, \Sigma) = \frac{1}{2\pi|det\,\Sigma|^{1/2}} \exp\left(\frac{1}{2}(\vec{q} - \vec{\mu})\Sigma^{-1}(\vec{q} - \vec{\mu})^T\right)$. Let $\Phi(\cdot; \vec{\mu}, \Sigma)$ denote the corresponding cumulative distribution function and let $A(\vec{a}, \vec{b}, \vec{\mu}, \Sigma) = \Phi(\vec{b}; \vec{\mu}, \Sigma) - \Phi([a_1, b_2]; \vec{\mu}, \Sigma) - \Phi([b_1, a_2]; \vec{\mu}, \Sigma) + \Phi(\vec{a}; \vec{\mu}, \Sigma)$. Set $f(\vec{q}; [\vec{a}, \vec{b}, \vec{\mu}, \Sigma]) = \frac{\phi(\vec{q}; \vec{\mu}, \Sigma)\chi_{[a_1,b_1]\times[a_2,b_2]}(\vec{q})}{A(\vec{a}, \vec{b}, \vec{\mu}, \Sigma)}$. In order to guarantee that we only search over positive definite symmetric matrices $\Sigma$, we parameterize $\Sigma$ as $\Sigma = LL^T$, where $L$ is lower triangular. It then follows that $\Sigma > 0$ so long as $L_{11}$ and $L_{22}$ are both nonzero. Thus the optimization is over a feasible (e.g., since our model is diffusion based, we would want $a_1, a_2 > 0$ or $\vec{\alpha} > 0$) subset of $\mathbb{R}^9$ with $\rho = [a_1, b_1, a_2, b_2, \mu_1, \mu_2, L_{11}, L_{21}, L_{22}]$.

All computations were in Matlab on either MAC or PC laptops or desktops. For higher dimensional problems with high resolution discretization, faster platforms (e.g. a cluster) may be required. The optimization problems ($\mathbf{P_N}$) were solved iteratively using FMINCON in the Matlab Optimization Toolbox. We computed gradients using the adjoint (see Sirlanci et al. (2019)); we also let FMINCON compute them using finite differences. Both yielded the same results. The finite difference calculations were faster, but the adjoint would likely be preferable for problems involving a higher dimensional parameter space such as would be encountered in the non-parametric case. This is because the required number of integrations of the state equation when using the adjoint method does not increase with the number of parameters to be estimated as it does with a finite difference scheme for computing the gradient of the cost functional. Initial estimates for the parameters were obtained by first fitting each dataset deterministically via nonlinear least squares to obtain estimates for $q_1$ and $q_2$. Then sample means, standard deviations, and covariances were used to compute initial estimates for $a_1, b_1, a_2, b_2, \mu_1, \mu_2$. The two random variables $q_1$ and $q_2$ were initially assumed to be independent, each with standard deviation one sixth of the length of the corresponding boundary of the initial domain. Some care must be exercised in choosing these initial guesses. Also, if in any iteration, the tails of the pdf become too close to zero (see Assumption (v)), the Galerkin scheme's mass and stiffness matrices can be close to singular.

A WRISTAS™7 (Fig. 1) was worn for 18 days by one of the authors (S.E.L.). It recorded TAC every 5-minutes. BrAC was measured every 30 min except during the first episode, in the lab, where it was measured every 15 min. In the field the sensor was worn and alcohol consumed ad libitum for 18 days. BrAC readings were taken every 30 min until they returned to 0.000. Fig. 2.a shows the 18-day TAC signal and the contemporaneous BrAC measurements. Eleven drinking episodes are marked. The TAC is in units of mg/dL, and the BrAC is in units of percent alcohol.

We first visually stratified the population into two groups, one containing the seven episodes in which the peak BrAC was higher than the (bench calibrated) peak TAC (episodes 1, 2, 4, 6, 7, 8, and 11) and a second containing the remaining four drinking episodes in which the reverse was true (episodes 3, 5, 9, and 10). Our results for the first stratified group are shown in Figs. 2.b–2.i. In Fig. 2.c, .d, .f, .g, and .i we plotted the training BrAC and TAC data for each of the episodes 1, 2, 6, 7, and 11 along with resulting fit population model estimated TAC and the 75% credible band. In Fig. 2.e and 2.h we plotted the results of a cross validation on episodes 4 and 8. In Fig. 2.b we plotted the optimal fit population truncated bivariate normal pdf. The converged values for the parameters were $a_1^* = 0.0000$, $b_1^* = 1.4850$, $a_2^* = 0.0000$, $b_2^* = 2.0363$, $\mu_1^* = 0.6318$, $\mu_2^* = 1.0295$, and $\Sigma_{1,1}^* = 0.0259$, $\Sigma_{2,2}^* = 0.1232$, and $\Sigma_{1,2}^* = \Sigma_{2,1}^* = 0.0077$.

The credible bands were computed using samples of $\vec{q} = [q_1, q_2]$ based on the optimal distribution using importance sampling and the state via $\langle c(\vec{q}), x_{i,j}^N\rangle_{V^*,V} = x_{i,j}^N(0, \vec{q})$, where $x_{i,j}^N = x_{i,j}^N(\eta, \vec{q})$, is given by (4.1). Strictly speaking, since $x_{i,j}^N \in \mathscr{V} = L_\pi^2(Q; V)$, pointwise evaluation is undefined. The results being reasonable and extremely useful; we nevertheless included them. Our results for all 11 episodes, for the second stratified group, and for a population with multiple subjects each with a single drinking episode wearing the SCRAM sensor (Fig. 1) were all similar to those presented above (Sirlanci et al., 2018).

## 7. Discussion and concluding remarks

We are investigating elimination of the requirement that the measures $\pi$ be defined in terms of a density. We believe that it is possible to directly apply the Prohorov metric based framework developed in Banks and Thompson (2012) by using a different version of the Trotter Kato-like semigroup approximation result in Section 4. We also believe that results for the estimation of functional parameters in parabolic systems could be used to estimate the pdfs non-parametrically. Indeed, in the system (3.1)–(3.2), the pdf effectively plays the role of a non-constant coefficient in an abstract parabolic system. Thus, we should be able to estimate both the support and the shape of the density by parameterizing the pdf as a linear combination of basis elements (e.g. splines, orthogonal polynomials, etc.) and then estimating the coefficients in the expansion. We are also looking at polynomial chaos expansions for $\vec{q}$ and then estimating the coefficients.

Ultimately we want to estimate the input, $u$, or the BAC/BrAC, from the output $y$, or TAC. Once the distribution of $q$ has been estimated, this takes the form of a deconvolution. In Sirlanci et al. (2018) we use the results presented here together with the framework in Gittelson et al. (2014) and Schwab and Gittelson (2011) to do just that. We obtain an estimate of BAC/BrAC along with credible bands. We are also looking at using the approach in Gittelson et al. (2014) and Schwab and Gittelson (2011) to control random parabolic systems, in particular, the computation of the feedback solution to the LQR and LQG problems.

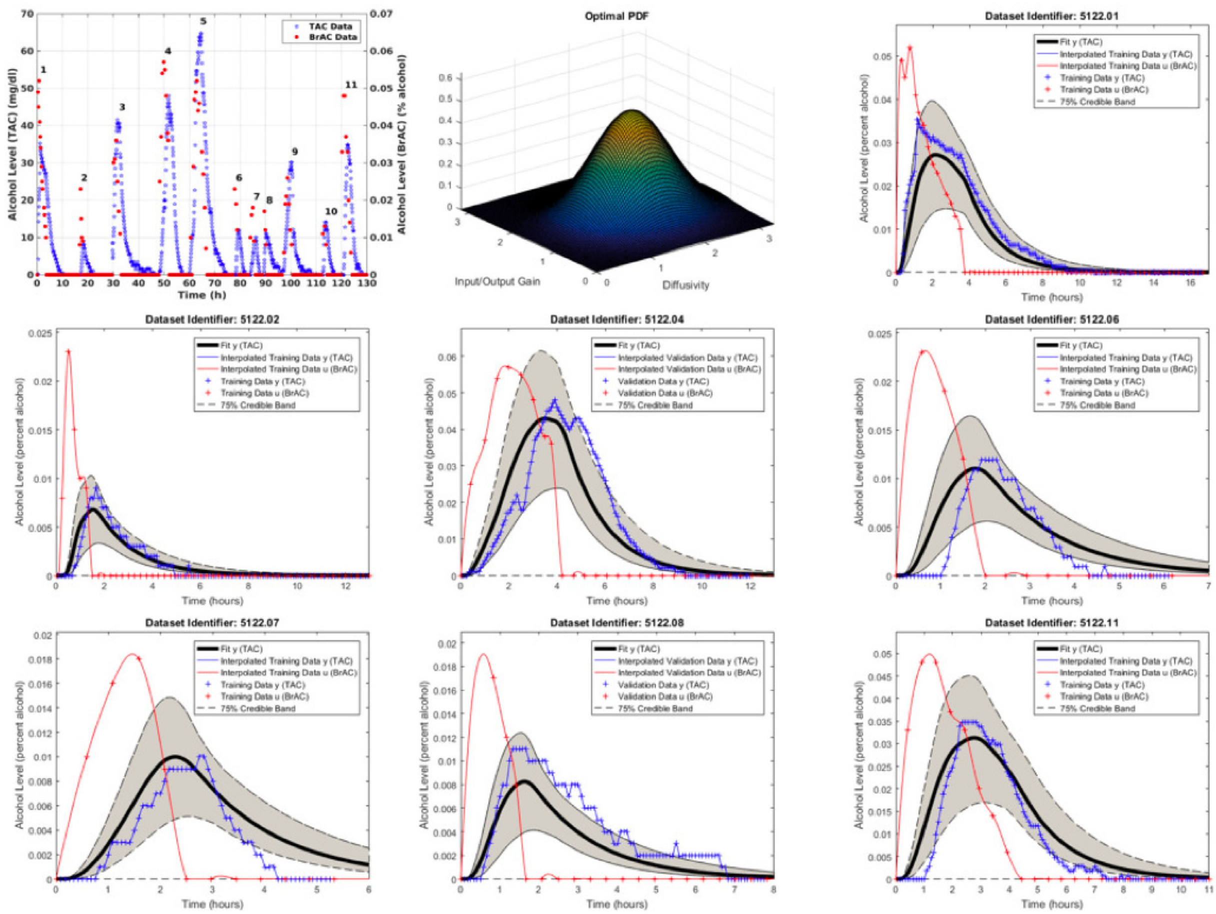



**Fig. 2.** Row-wise, left to right, from the top left: Panel (a) BrAC and TAC data for the 11 drinking episodes, (b) the optimal pdf, (c), (d), (f), (g), and (i) are the plots of training BrAC and corresponding TAC, resulting fit population model estimated TAC, and 75% credible band; (e) and (h) are the cross validation results.

## References


Banks, H. T., Burns, J. A., & Cliff, E. M. (1981). Parameter estimation and identification for systems with delays. *SIAM Journal on Control and Optimization*, *19*(6), 791–828.

Banks, H. T., Flores, K. B., Rosen, I. G., Rutter, E. M., Sirlanci, Melike, & Thompson, Clayton (2018). The prohorov metric framework and aggregate data inverse problems for random PDEs. *Communications in Applied Analysis*, *22*(3), 415–446.

Banks, H. T., Hu, Shuhua, & Thompson, W. Clayton (2014). *Monographs and research notes in mathematics*, *Modeling and inverse problems in the presence of uncertainty*. CRC Press, Chapman Hall.

Banks, H. T., & Ito, K. (1988). A unified framework for approximation in inverse problems for distributed parameter systems. *Control-theory and Advanced Technology*, *4*(1), 73–90.

Banks, H. T., & Kunisch, Karl (2012). *Estimation techniques for distributed parameter systems*. Springer Verlag.

Banks, H. T., & Thompson, W. Clayton (2012). *Least squares estimation of probability measures in the Prohorov metric framework*: *Technical report*, DTIC Document.

Calvetti, Daniela, Kaipio, Jario P., & Somersalo, Erkki (2014). Inverse problems in the Bayesian framework. *Inverse Problems*, *30*, 1–4.

Casella, George, & Berger, Roger L. (2002). *Statistical inference* (2nd ed.). Pacific Grove, CA: Duxbury.

Dai, Zheng, Rosen, I. G., Wang, Chunming, Barnett, Nancy, & Luczak, Susan E. (2016). 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: MBE*, *13*(5), 911.

Dashti, Masoumeh, & Stuart, Andrew M. (2017). The Bayesian approach to inverse problems. In R. Ghanem, & et al. (Eds.), *Handbook of Uncertainty Quantification* (pp. 311–428). Springer International Publishing Switzerland.

Davidian, M., & Giltinan, D. (1995). *Nonlinear models for repeated measurement data*. New York: Chapman and Hall.

Davidian, M., & Giltinan, D. M. (2003). Nonlinear models for repeated measurement data: An overview and update. *Journal of Agricultural, Biological, and Environmental Statistics*, *8*, 387–419.

Demidenko, E. (2013). *Mixed models, theory and applications* (2nd ed.). Hoboken: John Wiley and Sons.

Gittelson, Claude Jeffrey, Andreev, Roman, & Schwab, Christoph (2014). Optimality of adaptive Galerkin methods for random parabolic partial differential equations. *Journal of Computational and Applied Mathematics*, *263*, 189–201.

Kato, Tosio (2013). *Perturbation theory for linear operators, vol. 132*. Springer Science & Business Media.

Levi, A. F. J., & Rosen, I. G. (2010). A novel formulation of the adjoint method in the optimal design of quantum electronic devices. *SIAM Journal on Control and Optimization*, *48*(5), 3191–3223.

Lions, J. L. (1971). *Optimal control of systems governed by partial differential equations*. Springer-Verlag.

Mould, D. R., & Upton, R. N. (2012). Basic concepts in population modeling, simulation, and model-based drug development. *CPT: Pharmacometrics and Systems Pharmacology*, *1*(9), 1–14.

Pazy, A. (1983). *Semigroups of linear operators and applications to partial differential equations*. Springer.

Rosen, I. G., Luczak, Susan E., & Weiss, Jordan (2014). 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*, 357–376.

Schultz, M. H. (1972). *Prentice-hall series in automatic computation*, *Spline snalysis*. Pearson Education, Limited.

Schwab, Christoph, & Gittelson, Claude Jeffrey (2011). Sparse tensor discretizations of high-dimensional parametric and stochastic PDEs. *Acta Numerica*, *20*, 291–467.

Sirlanci, Melike, Luczak, Susan E., & Rosen, I. G. (2019). Estimation of the distribution of random parameters in discrete time abstract parabolic systems with unbounded input and output: Approximation and convergence. *Communications in Applied Analysis*, *23*, 44.

Sirlanci, Melike, Rosen, I. G., Luczak, Susan E., Fairbairn, Catharine E., Bresin, Konrad, & Kang, Dayheon (2018). Deconvolving the input to random abstract parabolic systems; A population model-based approach to estimating blood/breath alcohol concentration from transdermal alcohol biosensor data. *Inverse Problems*, *34*(12), arXiv:1807.05088v1.

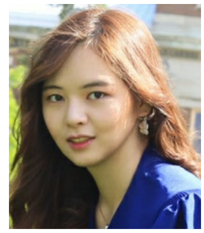
Stuart, A. M. (2010). Inverse problems: A Bayesian perspective. *Acta Numerica*, 451–559.

Tanabe, H. (1979). *Equations of evolution*. Pitman.

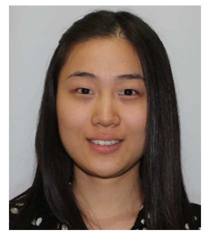
**Melike Sirlanci** received her B.Sc. degree in Mathematics from Dokuz Eylul University, Izmir, Turkey, in 2011 and her M.Sc. degree in Mathematics from Koc University, Istanbul, Turkey, in 2013. She received her Ph.D. degree in Applied Mathematics from the University of Southern California in 2018 along with a M.Sc. degree in Statistics. Currently, she is a postdoctoral researcher in the Department of Computing and Mathematical Sciences at the California Institute of Technology in Pasadena, California. Her research interests are primarily in the fields of modeling, estimation, control, and numerical optimization of systems governed by various types of differential equations and that are motivated by problems arising in the areas of biomedicine and mathematical biology.

**Susan E. Luczak** is a Research Professor in the Department of Psychology at the University of Southern California. Her research interests include genetic and ethnic group differences in the lifecourse development of alcohol use and problems, real-time drinking behavior, and biosensor alcoholic beverage consumption monitoring. In particular, she examines a broad range of biopsychosocial risk and protective factors associated with the development and transmission of alcohol-related behaviors in a variety of ethnic groups and cultural settings. Dr. Luczak received her Sc.B. in Psychology at Brown University in 1993 and her M.A. and Ph.D. in Clinical Psychology at the University of Southern California in 1997 and 2000, respectively. Dr. Luczak was a postdoctoral fellow at the University of California, San Diego, during the 2000–2001 academic year. Her research has been primarily funded by the National Institutes of Health (NIAAA, NIDA), as well as the Alcoholic Beverage Medical Research Foundation and Tobacco-Related Disease Research Program.

**Catharine E. Fairbairn** earned her Ph.D. in clinical and health psychology from the University of Pittsburgh in 2015. She began working with individuals with alcohol problems in 2004 in New York City, and has since been consistently working to build the understanding of addictions through her research and applied work at major research institutions across the country and around the world. She is currently an assistant professor of clinical psychology at the University of Illinois at Urbana-Champaign. Her research takes a multi-method approach to exploring the role of basic social and affective processes in alcohol use.



**Dahyeon Kang** received her B.S. from the University of Illinois at Urbana-Champaign in 2017 and is currently a graduate student in the Alcohol Research Laboratory in the Department of Psychology at UIUC. Her research focuses on employing electroencephalogram methods in alcohol administration paradigms to investigate the effects of alcohol on individuals' emotions and cognitions.



**Ruoxi Pan** obtained her bachelor's degree from the University of Southern California, in 2017. She served as an undergraduate student researcher in the Modeling and Simulation Laboratory in the USC Department of Mathematics in 2017. Her areas of interest include artificial intelligence, machine learning, and data science.

**Xin Yu** obtained her bachelor's degrees in Electrical Engineering and in Applied and Computational Mathematics from the University of Southern California, in 2017. She is currently a graduate student in mechanical and electrical engineering and in computer science and data science and systems at the University of California, Berkeley. She served as an undergraduate student researcher in the Modeling and Simulation Laboratory in the USC Department of Mathematics in 2017. Her areas of interest include machine learning, and data science.

**I. Gary Rosen** Gary Rosen's research is focused on topics related to the modeling, estimation, control, and optimal design of systems governed by infinite dimensional dynamical systems, and in particular, partial differential equations. He received his Sc.B. in Mathematics and his Sc.M. and Ph.D. degrees in Applied Mathematics, all from Brown University in providence, Rhode Island. He has been on the faculty of the Department of Mathematics at the University of Southern California since 1984 and served as the department chair from 2005–2011. He also served on the faculty of Bowdoin College and was a member of the Technical Staff at the Charles Stark Draper Laboratory in Cambridge, Massachusetts. Professor Rosen's research has been supported by grants from AFOSR, ONR, DARPA, NASA, NSF, and NIH, and has served as a member of the editorial board of the IEEE Transactions on Control Technology.