



Research article

Computation of nonparametric, mixed effects, maximum likelihood, biosensor data based-estimators for the distributions of random parameters in an abstract parabolic model for the transdermal transport of alcohol

Lernik Asserian¹, Susan E. Luczak² and I. G. Rosen^{3,*}

¹ Department of Mathematics, Stanford University, Stanford, CA 94305, USA

² Department of Psychology, University of Southern California, Los Angeles, CA 90089, USA

³ Department of Mathematics, University of Southern California, Los Angeles, CA 90089, USA

* **Correspondence:** Email: grosen@math.usc.edu; Tel: +13107102235.

Abstract: The existence and consistency of a maximum likelihood estimator for the joint probability distribution of random parameters in discrete-time abstract parabolic systems was established by taking a nonparametric approach in the context of a mixed effects statistical model using a Prohorov metric framework on a set of feasible measures. A theoretical convergence result for a finite dimensional approximation scheme for computing the maximum likelihood estimator was also established and the efficacy of the approach was demonstrated by applying the scheme to the transdermal transport of alcohol modeled by a random parabolic partial differential equation (PDE). Numerical studies included show that the maximum likelihood estimator is statistically consistent, demonstrated by the convergence of the estimated distribution to the “true” distribution in an example involving simulated data. The algorithm developed was then applied to two datasets collected using two different transdermal alcohol biosensors. Using the leave-one-out cross-validation (LOOCV) method, we found an estimate for the distribution of the random parameters based on a training set. The input from a test drinking episode was then used to quantify the uncertainty propagated from the random parameters to the output of the model in the form of a 95% error band surrounding the estimated output signal.

Keywords: nonparametric estimation; mixed effects model; maximum likelihood estimation; existence and consistency; random discrete-time dynamical systems; random partial differential equations; finite dimensional approximation and convergence; alcohol biosensor; transdermal alcohol concentration

1. Introduction

In clinical therapy, medical research, and law enforcement, the breathalyzer, developed by Borkenstein based on a redox reaction and Henry's law [1], is used to measure breath alcohol concentration (BrAC), a surrogate for blood alcohol concentration (BAC). Clinicians and researchers consider it to be reasonably accurate to substitute BrAC for BAC, and in general this continues to be the case across different environmental conditions and across different individuals [1]. Nevertheless, collecting near-continuous BrAC samples accurately (i.e., obtaining a deep lung sample that is not contaminated by any existing alcohol remaining in the mouth) is challenging and often impractical in the field.

Most of the ethanol, the type of alcohol in alcoholic beverages, that enters the human body, is metabolized by the liver into other products that are then excreted. In addition, a portion of ingested ethanol exits the body directly through exhalation and urination [2] and approximately 1% diffuses through the epidermal layer of the skin in the form of perspiration and sweat. The amount of alcohol excreted in this manner is quantified in the form of transdermal alcohol concentration (TAC). TAC has been shown to be largely positively correlated with BrAC and BAC [3]. However, the precise relationship between TAC and BrAC/BAC is complicated due to a number of confounding physiological, technological and environmental factors including, but not limited to, the skin's epidermal layer thickness, porosity and tortuosity, the process of vasodilation as observed through blood pressure and flow rate, the underlying technology of the particular sensor being used and the ambient temperature and humidity.

Currently, there are a number of different biosensors based on a variety of analog principles that can measure TAC essentially continuously, passively, unobtrusively, and relatively accurately, and make it available for processing in real time. Some of these devices are already commercially available and more are on the way. Several of these biosensors, like the breathalyzer, rely on relatively standard fuel-cell technology (i.e., converting chemical energy into electricity through redox reactions) to effectively count the number of ethanol molecules that evaporate during perspiration from the epidermal layer of the skin in near-continuous time [4]. Figure 1 shows two of these TAC measuring devices: The WrisTASTM7 developed by Giner, Inc. in Waltham, MA, and the SCRAM CAM[®] (Secure Continuous Remote Alcohol Monitor) developed by Alcohol Monitoring Systems, Inc. (AMS) in Littleton, Colorado.



Figure 1. WrisTASTM7 (left) and SCRAM CAM[®] (right) transdermal alcohol biosensors.

Historically, researchers, clinicians, and the courts have always relied on BrAC or, when available, BAC. Consequently, in order to make TAC biosensors practical and accepted by the alcohol

community, reliable and consistent means for converting TAC into equivalent BrAC or BAC must be developed, and this involves challenges that must be dealt with as indicated previously. In the past, our approach to developing a method for converting TAC into BrAC or BAC was based on deterministic methods for estimating parameters in distributed parameter systems, such as those described in [5, 6]. Our earlier work along these lines has been reported in, for example, [7–9]. In these treatments, a forward model in the form of a one-dimensional diffusion equation based on Fick's law [10] with BrAC as the input and TAC as the output was first calibrated (i.e., fit) using BrAC and TAC data collected from the patient or research subject in the clinic or laboratory during what is known as a controlled alcohol challenge. Then, after the same patient or research subject has worn the TAC sensor in the field for an extended period of time (e.g., days, weeks, or even months), the TAC data is downloaded, and the fit forward model is used to deconvolve the BrAC or BAC input from the observed TAC output.

In order to eliminate the calibration process, we developed a population model-based approach wherein the parameters in the model were assumed to be random. Then, rather than fitting the values of the parameters themselves, their distributions were estimated based on BrAC and TAC data from a cohort of individuals (see, for example, [11–17]). We have also developed a number of physics-informed (based on the first principles diffusion-based model given in (1.1)–(1.5) below) data-based machine learning schemes using hidden Markov models, generative adversarial and long short-term memory neural network models for estimating BrAC from biosensor TAC data, which are also trained using population data and do not require individual calibration [18–20].

In all of our approaches to the TAC to BAC/BrAC conversion problem, the underlying model was taken to be based on the first principles physics based initial-boundary value problem for a parabolic partial differential equation. This will also be the basic model to which we will direct our efforts in the present treatment. Transformed to be in terms of dimensionless variables, the model is given by

$$\frac{\partial x}{\partial t}(t, \eta) = q_1 \frac{\partial^2 x}{\partial \eta^2}(t, \eta), \quad 0 < \eta < 1, \quad t > 0, \quad (1.1)$$

$$q_1 \frac{\partial x}{\partial \eta}(t, 0) = x(t, 0), \quad t > 0, \quad (1.2)$$

$$q_1 \frac{\partial x}{\partial \eta}(t, 1) = q_2 u(t), \quad t > 0, \quad (1.3)$$

$$x(0, \eta) = x_0, \quad 0 < \eta < 1, \quad (1.4)$$

$$y(t) = x(t, 0), \quad t > 0, \quad (1.5)$$

where t and η are the temporal and spatial variables, respectively, and $x(t, \eta)$ indicates the concentration of ethanol in the epidermal layer of the skin at time t and depth η , where $\eta = 0$ is at the skin surface and $\eta = 1$ is at the boundary between the epidermal and dermal layers of the skin. The input to the system is $u(t)$, which is the BrAC/BAC at time t , and the output is $y(t)$, which is the TAC at time t . Equation (1.1) represents the transport of ethanol through the epidermal layer of the skin. The boundary conditions (1.2) and (1.3) represent respectively the evaporation of ethanol at the skin surface and the flux of ethanol across the boundary between the epidermal and dermal layers. It is assumed that there is no alcohol in the epidermal layer of the skin at time $t = 0$, so the initial condition (1.4) is $x_0(\eta) = 0$, $0 < \eta < 1$. Finally, the output Eq (1.5) represents the TAC level measured by the biosensor at the skin surface.

The parameters in the system (1.1)–(1.5) that will be assumed to be random are q_1 and q_2 , which represent respectively the normalized diffusivity and the normalized flux gain at the boundary between the dermal and epidermal layers. The values or distributions of these parameters are assumed to depend on environmental conditions, the particular sensor being used, and the physiological characteristics of the individual wearing the sensor. The parameter vector is $\mathbf{q} = (q_1, q_2) \in Q$, where Q is assumed to be a compact subset of $\mathbb{R}^+ \times \mathbb{R}^+$ with metric d .

In population modeling, we can statistically classify the methods as parametric or nonparametric. In the parametric approach, we assume that the general structure of the distribution is known a-priori but with unknown parameters. Then, for example, if we know that the distribution is normal with unknown mean and variance, the estimation problem is to estimate these two unknown parameters. On the other hand; in the nonparametric approach, the structure of the distribution is assumed to be unknown, and the problem is to estimate the distribution itself. In either of these paradigms, different statistical approaches to the estimation problem can be taken. For example, in [15–17], a parametric least squares naive pooled data approach was used, while in [21, 22], the approach was nonparametric. In [12], a Bayesian framework was developed, and in the present treatment we consider a mixed effects (see, for example, [23–25]) maximum likelihood-based statistical model. In the mixed effects model, it is assumed that observations are specific to a single individual plus a random error. The mixed effects model is a combination of the fixed-effects model, which describes the characteristics for an average individual in the population, and the random-effects model, which describes the inter-individual variability [26]. An overview of these different statistical approaches in the context of pharmacokinetics can be found in [27].

While TAC and BrAC are related, they are not precisely equivalent in terms of quantification. BrAC serves as a reliable indicator of BAC, offering valuable quantitative insights into the immediate impact of alcohol on judgment, motor skills and cognition. On the contrary, TAC lags behind BrAC because it must pass through the skin before measurement. TAC levels exhibit variability among individuals, devices and across drinking episodes, in contrast to the relative consistency of BrAC across different people and conditions. Consequently, interpreting TAC in relation to real-time blood alcohol levels becomes challenging, limiting its usefulness as a quantitative alcohol measure. Therefore, having a method to estimate BrAC from TAC would enhance its practicality for researchers, healthcare professionals and individuals seeking meaningful and time-sensitive information about their alcohol levels. The inverse problem of estimating the input BrAC signal u to the system (1.1)–(1.5) based on observations of the TAC output signal y is the central focus of our team's research effort. In our treatment here, we consider only the problem of estimating the distribution of the random parameters in the population model and how the uncertainty in the parameters propagates through the system of Eqs (1.1)–(1.5) to the TAC output y . Once the distributions of the random parameters in the the model (1.1)–(1.5) have been estimated, we have previously developed (and continue to investigate) a number of deconvolution or inverse filtering methods that can then be used to obtain an estimate for the BrAC signal along with a credible or error band that result from the uncertainty in the model parameters as described by their estimated distributions. We have looked at a number of approaches including frequency domain, Bayesian, linear-quadratic Gaussian compensator, maximum likelihood, autoregressive moving average (ARMA) and nonlinear least squares based techniques [11, 28–30].

In addition to the work of our group on TAC to BAC/BrAC conversion cited above, other researchers

have also been looking at this problem and have tried a number of different approaches. For example, in [31, 32], a more traditional approach based on standard linear regression techniques was developed and discussed. A number of ideas from the machine learning literature have also been considered. In [33], a scheme based on random forests was used to recover BrAC from TAC, and in our group, there are the machine learning based treatments cited earlier [18–20].

An outline of the remainder of the paper is as follows. In Section 2, we provide a summary of the Prohorov metric on the set of probability measures as it was used by Banks and his coauthors in [21]. In Section 3, we define our mathematical model in the form of a random discrete-time dynamical system and we define the maximum likelihood estimator for the distribution of the random parameters. In Section 4, we establish the existence and consistency of the maximum likelihood estimator for the distribution of the random parameters, while in Section 5, we demonstrate the convergence of finite dimensional approximations for our estimator. In Section 6, we summarize results for abstract parabolic systems, their finite dimensional approximation and an associated convergence theory. In Section 7, the application of our scheme to the transdermal transport of alcohol is presented and discussed. This includes numerical studies for two examples, one involving simulated data and the other actual data collected in the laboratory of one of the coauthors, Dr. Susan Luczak in the Department of Psychology at University of Southern California (USC). For the simulated data example in Section 7.1, we are able to observe the convergence of the estimated distribution of the random parameter vector $\mathbf{q} = (q_1, q_2)$ to the “true” distribution as the number of drinking episodes increases, as the number of Dirac measure nodes increases and as the level of discretization in the finite dimensional approximations increases. In the actual data example discussed in Section 7.2, we apply the leave-one-out cross-validation (LOOCV) method by first estimating the distribution of the parameter vector \mathbf{q} using a training set, and then estimating the TAC output using the estimated distribution and the BrAC input of a testing episode.

2. Prohorov metric framework

Banks and his coauthors developed a framework for estimation of the probability measure for random parameters in continuous-time dynamical systems based on the Prohorov metric [21]. Here, we summarize the Prohorov metric and its properties.

Let Q be a Hausdorff metric space with metric d . Define

$$C_b(Q) = \{f : Q \rightarrow \mathbb{R} \mid f \text{ is bounded and continuous}\},$$

and given any probability measure $P \in \mathcal{P}(Q)$, where $\mathcal{P}(Q)$ denotes the set of all probability measures defined on Σ_Q , the Borel sigma algebra on Q , and some $\varepsilon > 0$, an ε -neighborhood of P is defined by

$$B_\varepsilon(P) = \left\{ \tilde{P} \mid \left| \int_Q f(\mathbf{q}) d\tilde{P}(\mathbf{q}) - \int_Q f(\mathbf{q}) dP(\mathbf{q}) \right| < \varepsilon, \text{ for all } f \in C_b(Q) \right\}.$$

Let $E \in \Sigma_Q$, and define the ε -neighborhood of E by

$$E^\varepsilon = \{\tilde{\mathbf{q}} \in Q \mid d(\tilde{\mathbf{q}}, E) < \varepsilon\} = \{\tilde{\mathbf{q}} \in Q \mid \inf_{\mathbf{q} \in E} d(\mathbf{q}, \tilde{\mathbf{q}}) < \varepsilon\}.$$

Given two probability measures, P and \tilde{P} in $\mathcal{P}(Q)$, the Prohorov metric ρ on $\mathcal{P}(Q) \times \mathcal{P}(Q)$ is defined such that

$$\tilde{P} \in B_\varepsilon(P) \iff \rho(P, \tilde{P}) < \varepsilon,$$

where

$$\rho(P, \tilde{P}) = \inf\{\varepsilon > 0 \mid \tilde{P}(E) \leq P(E^\varepsilon) + \varepsilon \text{ and } P(E) \leq \tilde{P}(E^\varepsilon) + \varepsilon, \text{ for all } E \in \Sigma_Q\}.$$

It can be shown that $(\mathcal{P}(Q), \rho)$ is a metric space. Also, the Prohorov metric metrizes the weak convergence of measures, i.e., given a sequence of measures $P_M \in \mathcal{P}(Q)$, for all $M = 1, 2, \dots$, and $P \in \mathcal{P}(Q)$,

$$P_M \xrightarrow{w^*} P \iff \rho(P_M, P) \rightarrow 0.$$

It is important to note that the weak* topology and the weak topology are equivalent on the space of probability measures.

For some n_q , $Q \subseteq \mathbb{R}^{n_q}$ and $P \in \mathcal{P}(Q)$, consider the random vector $X : Q \rightarrow \mathbb{R}^{n_q}$ on the probability space (Q, Σ_Q, P) given by $X(\mathbf{q}) = \mathbf{q}$ for $\mathbf{q} \in Q$. The cumulative distribution function for X is given by $F_X(q_1, \dots, q_n) = P(X \in \times_{\ell=1}^{n_q} (-\infty, q_\ell]) = P(\times_{\ell=1}^{n_q} \{(-\infty, q_\ell] \cap Q\})$. In this case, it follows that if $P_M, P_0 \in (\mathcal{P}(Q), \rho)$, for $M = 1, 2, \dots$, then $\rho(P_M, P_0) \rightarrow 0$ if and only if $F_{X_M} \rightarrow F_{X_0}$ at all points of continuity of F_{X_0} . Consequently, Prohorov metric convergence and weak and weak* convergence in $\mathcal{P}(Q)$ are also referred to as convergence in distribution.

If $\mathbf{q}_1, \mathbf{q}_2 \in Q$, then $\rho(\delta_{\mathbf{q}_1}, \delta_{\mathbf{q}_2}) = \min\{d(\mathbf{q}_1, \mathbf{q}_2), 1\}$, where $\delta_{\mathbf{q}_j} \in D = \{\delta_{\mathbf{q}} \mid \mathbf{q} \in Q\}$, the space of Dirac measures on Q , where for all $E \in \Sigma_Q$,

$$\delta_{\mathbf{q}}(E) = \begin{cases} 1 & \text{if } \mathbf{q} \in E \\ 0 & \text{if } \mathbf{q} \notin E \end{cases}.$$

The metric space $(\mathcal{P}(Q), \rho)$ is separable if and only if the metric space (Q, d) is separable. The sequence $\{\mathbf{q}_j\}_{j=1}^\infty$ is Cauchy in (Q, d) if and only if the sequence $\{\delta_{\mathbf{q}_j}\}_{j=1}^\infty$ is Cauchy in $(\mathcal{P}(Q), \rho)$. We also have that (Q, d) is complete if and only if $(\mathcal{P}(Q), \rho)$ is complete, and (Q, d) is compact if and only if $(\mathcal{P}(Q), \rho)$ is compact. The details and proofs can be found in [21].

Assume the metric space (Q, d) is separable and let $Q_d = \{\mathbf{q}_j\}_{j=1}^\infty$ be a countable dense subset of Q . Define the dense (see [21]) subset of $\mathcal{P}(Q)$, $\tilde{\mathcal{P}}_d(Q)$, as

$$\tilde{\mathcal{P}}_d(Q) = \{P \in \mathcal{P}(Q) \mid P = \sum_{j=1}^M p_j \delta_{\mathbf{q}_j}, \mathbf{q}_j \in Q_d, M \in \mathbb{N}, p_j \in [0, 1] \cap \mathbb{Q}, \sum_{j=1}^M p_j = 1\}, \quad (2.1)$$

the collection of all convex combinations of Dirac measures on Q with rational weights p_j at nodes $\mathbf{q}_j \in Q_d$, and for each $M \in \mathbb{N}$ let

$$\mathcal{P}_M(Q) = \{P \in \tilde{\mathcal{P}}_d(Q) \mid P = \sum_{j=1}^M p_j \delta_{\mathbf{q}_j}, \mathbf{q}_j \in \{\mathbf{q}_j\}_{j=1}^M\}.$$

3. The mathematical model

Consider the following discrete-time mathematical model for the i^{th} subject at time-step k

$$\begin{aligned}x_{k,i}(\mathbf{q}_i) &= g_{k-1}(x_{k-1,i}(\mathbf{q}_i), u_{k-1,i}; \mathbf{q}_i), \quad k = 1, \dots, n_i, \quad i = 1, \dots, m, \\x_{0,i} &= \phi_{0,i}, \quad i = 1, \dots, m,\end{aligned}$$

where \mathbf{q}_i is the i^{th} subject's parameter vector in Q , denoting the set of admissible parameters, $g_{k-1} : \mathcal{H} \times \mathbb{R}^V \times Q \rightarrow \mathcal{H}$, \mathcal{H} is in general an infinite dimensional Hilbert space, and $u_{k-1,i} \in \mathbb{R}^V$ is the input. The output is given by

$$y_{k,i}(\mathbf{q}_i) = h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i), \quad k = 1, \dots, n_i, \quad i = 1, \dots, m,$$

where $h_k : \mathcal{H} \times \mathcal{H} \times \mathbb{R}^V \times Q \rightarrow \mathbb{R}$.

For the mixed effects model, we define

$$Y_{k,i} = y_{k,i}(\mathbf{q}_i) + e_{k,i}, \quad k = 1, \dots, n_i, \quad i = 1, \dots, m, \quad (3.1)$$

where for each $i = 1, 2, \dots, m$, $e_{k,i}$ are independent and identically distributed (i.i.d.) with mean zero, variance σ^2 and $e_{k,i} \sim \varphi$, $k = 1, 2, \dots, n_i$, where φ is a density with respect to a sigma finite measure μ on \mathbb{R} and assumed to be continuous on \mathbb{R} . We assume that the random vectors $[e_{1,i}, \dots, e_{n_i,i}]$ are independent with respect to i , $i = 1, 2, \dots, m$; that is, the error is independent across individuals and conditionally independent within individuals (i.e., given \mathbf{q}_i). For each $i = 1, 2, \dots, m$, let $\mathbf{Y}_i = [Y_{1,i}, Y_{2,i}, \dots, Y_{n_i,i}]^T$, $\mathbf{y}_i(\mathbf{q}_i) = [y_{1,i}(\mathbf{q}_i), y_{2,i}(\mathbf{q}_i), \dots, y_{n_i,i}(\mathbf{q}_i)]^T$, $\mathbf{e}_i = [e_{1,i}, e_{2,i}, \dots, e_{n_i,i}]^T$ and rewrite (3.1) as

$$\mathbf{Y}_i = \mathbf{y}_i(\mathbf{q}_i) + \mathbf{e}_i, \quad i = 1, \dots, m.$$

Then, for $i = 1, 2, \dots, m$, \mathbf{Y}_i are independent with $\mathbf{Y}_i \sim f_i(\cdot; \mathbf{q}_i) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$, where

$$f_i(\mathbf{v}; \mathbf{q}_i) = \prod_{k=1}^{n_i} \varphi(v_k - y_{k,i}(\mathbf{q}_i)), \quad i = 1, \dots, m, \quad (3.2)$$

where $\mathbf{v} = [v_1, v_2, \dots, v_{n_i}]^T \in \mathbb{R}^{n_i}$. Let $P \in \mathcal{P}(Q)$ denote a probability measure on Σ_Q , where $\mathcal{P}(Q)$ denotes the set of all probability measures defined on Σ_Q , and let $P_0 \in \mathcal{P}(Q)$ be the "true" distribution of the random vector \mathbf{q}_i . The goal is to find an estimate of P_0 . In order to generate an estimator for P_0 and establish theoretical results and computational tools, we use the nonparametric maximum likelihood (NPML) approach introduced by Lindsay and Mallet in [34, 35], as well as the Prohorov metric-based framework on $\mathcal{P}(Q)$ introduced by Banks and his coauthors in [21], summarized in Section 2.

For $P \in \mathcal{P}(Q)$ and $i = 1, 2, \dots, m$, let

$$\begin{aligned}\mathcal{L}_i(P; \mathbf{Y}_i) &= \int_Q f_i(\mathbf{Y}_i; \mathbf{q}_i) dP(\mathbf{q}_i) \\ &= \int_Q \prod_{k=1}^{n_i} \varphi(Y_{k,i} - h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i)) dP(\mathbf{q}_i)\end{aligned} \quad (3.3)$$

be the contribution of the i^{th} subject to the likelihood function

$$\begin{aligned}\mathcal{L}_{\mathbf{n},m}(P;\mathbf{Y}) &= \prod_{i=1}^m \mathcal{L}_i(P;\mathbf{Y}_i) \\ &= \prod_{i=1}^m \int_Q f_i(\mathbf{Y}_i;\mathbf{q}_i) dP(\mathbf{q}_i) \\ &= \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(Y_{k,i} - h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i)) dP(\mathbf{q}_i),\end{aligned}\quad (3.4)$$

where $\mathbf{n} = \{n_i\}_{i=1}^m$ and $\mathbf{Y} = \{\mathbf{Y}_i\}_{i=1}^m$. The goal is to find P that maximizes the likelihood function.

Define the estimator

$$\begin{aligned}P_{\mathbf{n},m} &= \arg \max_{P \in \mathcal{P}(Q)} \mathcal{L}_{\mathbf{n},m}(P;\mathbf{Y}) \\ &= \arg \max_{P \in \mathcal{P}(Q)} \prod_{i=1}^m \int_Q f_i(\mathbf{Y}_i;\mathbf{q}_i) dP(\mathbf{q}_i) \\ &= \arg \max_{P \in \mathcal{P}(Q)} \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(Y_{k,i} - h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i)) dP(\mathbf{q}_i).\end{aligned}\quad (3.5)$$

Let $\hat{y}_{k,i}$ be realizations of the random variables $Y_{k,i}$, and define

$$\begin{aligned}\hat{P}_{\mathbf{n},m} &= \arg \max_{P \in \mathcal{P}(Q)} \mathcal{L}_{\mathbf{n},m}(P;\hat{\mathbf{y}}) \\ &= \arg \max_{P \in \mathcal{P}(Q)} \prod_{i=1}^m \int_Q f_i(\hat{\mathbf{y}}_i;\mathbf{q}_i) dP(\mathbf{q}_i) \\ &= \arg \max_{P \in \mathcal{P}(Q)} \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i)) dP(\mathbf{q}_i),\end{aligned}\quad (3.6)$$

where $\hat{\mathbf{y}} = \{\hat{\mathbf{y}}_i\}_{i=1}^m$, with $\hat{\mathbf{y}}_i = [\hat{y}_{1,i}, \hat{y}_{2,i}, \dots, \hat{y}_{n_i,i}]^T$.

The results of Lindsay and Mallet in [34, 35] states that the maximum likelihood estimator $\hat{P}_{\mathbf{n},m}$ can be found in the class of discrete distributions with at most m support points; that is, $\hat{P}_{\mathbf{n},m} \in \mathcal{P}_M(Q)$, where $M \leq m$. So, we define our approximating estimator over the set $\mathcal{P}_M(Q)$ where M denotes the number of nodes, $\{\mathbf{q}_j\}_{j=1}^M$. As a result, the optimization is over a finite set of parameters, being the rational weights $\{p_j\}_{j=1}^M$.

Also, we cannot exactly compute the maximum likelihood estimator $\hat{P}_{\mathbf{n},m}$ since $y_{k,i}$ must be approximated numerically by $y_{k,i}^N$, using a Galerkin numerical scheme with N denoting the level of discretization. Thus, our approximating estimator is

$$\begin{aligned}\hat{P}_{\mathbf{n},m,M}^N &= \arg \max_{P \in \mathcal{P}_M(Q)} \mathcal{L}_{\mathbf{n},m}^N(P;\hat{\mathbf{y}}) \\ &= \arg \max_{P \in \mathcal{P}_M(Q)} \prod_{i=1}^m \int_Q f_i^N(\hat{\mathbf{y}}_i;\mathbf{q}_i) dP(\mathbf{q}_i) \\ &= \arg \max_{P \in \mathcal{P}_M(Q)} \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i)) dP(\mathbf{q}_i).\end{aligned}\quad (3.7)$$

4. Existence and consistency of the maximum likelihood estimator

In [34], the existence and uniqueness of a maximum likelihood estimator of a mixing distribution using the geometry of mixture likelihoods was established. Similarly, in [35], the existence and uniqueness of the maximum likelihood estimator for the distribution of the parameters of a random coefficient regression model was established. Here, we provide an existence argument based on the maximization of a continuous function over a compact set.

The following theorem establishes the existence of the estimator $\hat{P}_{n,m}$ in (3.7), obtained from the realizations $\{\hat{y}_{k,i}\}$, $k = 1, \dots, n_i$, $i = 1, \dots, m$ of the random variables $\{Y_{k,i}\}$, $k = 1, \dots, n_i$, $i = 1, \dots, m$. This is sufficient for establishing the existence of the maximum likelihood estimator $P_{n,m}$ in (3.5).

Theorem 4.1. For $i = 1, 2, \dots, m$, let \mathcal{L}_i be given by Eq (3.3) and let $\mathcal{L}_{n,m}(P; \hat{\boldsymbol{y}})$ be given by equation (3.4), where $\hat{\boldsymbol{y}} = \{\hat{\boldsymbol{y}}_i\}_{i=1}^m$ with $\hat{\boldsymbol{y}}_i = [\hat{y}_{1,i}, \hat{y}_{2,i}, \dots, \hat{y}_{n_i,i}]^T$. Assume that for each $\hat{\boldsymbol{y}}_i$, we have a continuous function $\mathcal{L}_i(\cdot; \hat{\boldsymbol{y}}_i) : \mathcal{P}(Q) \rightarrow \mathbb{R}$, and also for each $P \in \mathcal{P}(Q)$ we have a measurable function $\mathcal{L}_i(P; \cdot) : \mathbb{R}^{n_i} \rightarrow \mathbb{R}$. Then there exists a measurable function $\hat{P}_{n,m} : \prod_{i=1}^m \mathbb{R}^{n_i} \rightarrow \mathcal{P}(Q)$ such that

$$\mathcal{L}_{n,m}(\hat{P}_{n,m}; \hat{\boldsymbol{y}}) = \sup_{P \in \mathcal{P}(Q)} \mathcal{L}_{n,m}(P; \hat{\boldsymbol{y}}).$$

Proof. The theorem can be proven in a similar way as in [21] with the difference that we are taking the sup (instead of the inf) of a continuous function over a compact set.

In order to establish the consistency of the maximum likelihood estimator $P_{n,m}$, we show that $\rho(P_{n,m}, P_0)$ converges almost surely to zero. We do this by applying a theorem by Kiefer and Wolfowitz in [36], establishing that the nonparametric maximum likelihood approach is statistically consistent. In other words, as the number of subjects m gets larger, the estimator $P_{n,m}$ converges in probability to P_0 , the “true” distribution in the sense of the Prohorov metric, or weakly or in distribution. Here, we have set up our problem in a way that makes establishing the consistency a straightforward application of the consistency result in [36].

Theorem 4.2. For each $i = 1, \dots, m$, assume that the map $\mathbf{q}_i \mapsto f_i(\mathbf{v}; \mathbf{q}_i)$ from Q into \mathbb{R} is continuous for each $\mathbf{v} \in \mathbb{R}^{n_i}$, and $f_i(\mathbf{v}; \mathbf{q}_i)$ is measurable in \mathbf{v} for any $\mathbf{q}_i \in Q$, where f_i is given by (3.2). Assume further that P_0 is identifiable; that is, for $P_1 \in \mathcal{P}(Q)$ with $P_1 \neq P_0$, we have

$$\begin{aligned} & \prod_{i=1}^m \int_0^{z_i} \int_Q \prod_{k=1}^{n_i} \varphi(z_k - h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i)) dP_1(\mathbf{q}_i) d\mu^{n_i} \\ & \neq \prod_{i=1}^m \int_0^{z_i} \int_Q \prod_{k=1}^{n_i} \varphi(z_k - h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i)) dP_0(\mathbf{q}_i) d\mu^{n_i}, \end{aligned}$$

for at least one $\mathbf{z} = [z_1^T, \dots, z_m^T]^T \in \mathbb{R}^{\sum_{i=1}^m n_i}$, where for $i = 1, 2, \dots, m$, $\mathbf{z}_i = [z_{1,i}, \dots, z_{n_i,i}]^T \in \mathbb{R}^{n_i}$, and the technical integrability assumption holds; that is, for any $P \in \mathcal{P}(Q)$,

$$\lim_{\varepsilon \downarrow 0} E_{P_0} \left[\log \frac{\sup_{\tilde{P} \in B_\varepsilon(P)} \mathcal{L}_i(\tilde{P}; \mathbf{Y}_i)}{\mathcal{L}_i(P_0; \mathbf{Y}_i)} \right]_+ < \infty,$$

where \mathcal{L}_i is given by Eq (3.3). Then, as $m \rightarrow \infty$, $\rho(P_{n,m}, P_0) \rightarrow 0$ almost surely (i.e., with probability 1), and therefore in probability as well.

Proof. The assumptions we have made in the previous section and in the statement of the theorem are sufficient to argue that Assumptions 1–5 in [36] are satisfied. The conclusion of the consistency result in [36] is that the cumulative distribution functions $F_{n,m}$ corresponding to $P_{n,m}$ converge almost surely to the cumulative distribution function F_0 corresponding to P_0 at every point of continuity of F_0 . It follows that $\rho(P_{n,m}, P_0) \rightarrow 0$ almost surely (i.e., with probability 1); thus, in probability as well, as $m \rightarrow \infty$, and the theorem is proven.

5. Convergence of the finite dimensional approximations

We want to establish the convergence of the finite dimensional maximum likelihood estimators to the maximum likelihood estimator corresponding to the infinite dimensional model. As mentioned earlier, we cannot actually compute $\hat{P}_{n,m}$ in (3.6) and, consequently, we approximate it by $\hat{P}_{n,m,M}^N$ in (3.7). Consider the following assumptions:

A1. For all \mathbf{n} , m , and N , the map $P \mapsto \mathcal{L}_{n,m}^N(P; \hat{\boldsymbol{\psi}})$ is a continuous map.

A2. For any $P_M, P \in \mathcal{P}(Q)$, $M = 1, 2, \dots$, such that $\rho(P_M, P) \rightarrow 0$, we have $\mathcal{L}_i^N(P_M; \hat{\boldsymbol{\psi}}_i) \rightarrow \mathcal{L}_i(P; \hat{\boldsymbol{\psi}}_i)$ as $N, M \rightarrow \infty$ for $i = 1, \dots, m$.

A3. For all $P \in \mathcal{P}(Q)$ and $i = 1, \dots, m$, $\mathcal{L}_i^N(P_M; \hat{\boldsymbol{\psi}}_i)$ and $\mathcal{L}_i(P; \hat{\boldsymbol{\psi}}_i)$ are uniformly bounded.

Theorem 5.1. Under assumptions **A1–A3**, there exists maximizers $\hat{P}_{n,m,M}^N$ given by Eq (3.7). In addition, there exists a subsequence of $\hat{P}_{n,m,M}^N$ that converges to $\hat{P}_{n,m}$ given by Eq (3.6) as $M, N \rightarrow \infty$.

Proof. For all \mathbf{n} , m , and N , by continuity of the map $P \mapsto \mathcal{L}_{n,m}^N(P; \hat{\boldsymbol{\psi}})$ per assumption **A1** and compactness of $(\mathcal{P}(Q), \rho)$, we can conclude that $\hat{P}_{n,m,M}^N$ exists.

In [21], it is shown that $\tilde{\mathcal{P}}_d(Q)$ given by Eq (2.1) is a dense subset of $\mathcal{P}(Q)$. Thus, for $M = 1, 2, \dots$, construct a sequence of probability measures $P_M \in \mathcal{P}_M(Q) \subset \tilde{\mathcal{P}}_d(Q) \subset \mathcal{P}(Q)$, such that $\rho(P_M, P) \rightarrow 0$ in $\mathcal{P}(Q)$. Then, by assumptions **A2** and **A3**, we have

$$\left| \mathcal{L}_{n,m}^N(P_M; \hat{\boldsymbol{\psi}}) - \mathcal{L}_{n,m}(P; \hat{\boldsymbol{\psi}}) \right| = \left| \prod_{i=1}^m \mathcal{L}_i^N(P_M; \hat{\boldsymbol{\psi}}_i) - \prod_{i=1}^m \mathcal{L}_i(P; \hat{\boldsymbol{\psi}}_i) \right| \rightarrow 0.$$

Consequently, $\mathcal{L}_{n,m}^N(P_M; \hat{\boldsymbol{\psi}}) \rightarrow \mathcal{L}_{n,m}(P; \hat{\boldsymbol{\psi}})$ as $M, N \rightarrow \infty$.

In addition, by definition, for each \mathbf{n} , m and N and for all $P_M \in \mathcal{P}_M(Q)$, we have

$$\mathcal{L}_{n,m}^N(\hat{P}_{n,m,M}^N; \hat{\boldsymbol{\psi}}) \leq \mathcal{L}_{n,m}^N(P_M; \hat{\boldsymbol{\psi}}). \quad (5.1)$$

In addition, by compactness of $\mathcal{P}(Q)$, there exists a subsequence of $\hat{P}_{n,m,M}^N$ that converges to $\hat{P}_{n,m}$ as $M, N \rightarrow \infty$. Thus, by taking the limit in (5.1) as $M, N \rightarrow \infty$, for all $P \in \mathcal{P}(Q)$, we find that

$$\mathcal{L}_{n,m}(\hat{P}_{n,m}; \hat{\boldsymbol{\psi}}) \leq \mathcal{L}_{n,m}(P; \hat{\boldsymbol{\psi}});$$

thus, $\hat{P}_{n,m} = \arg \max_{P \in \mathcal{P}(Q)} \mathcal{L}_{n,m}(P; \hat{\boldsymbol{\psi}})$ as given in Eq (3.6).

In practice, to achieve a desired level of accuracy, M and N are fixed sufficiently large. We choose a sufficiently large value for N . How large that needs to be, of course, depends on the particular numerical discretization scheme chosen. The most common choice would be using a Galerkin-based method to approximate $y_{k,i}$ by $y_{k,i}^N$, where

$$y_{k,i}^N = h_k(x_{k,i}^N(\mathbf{q}_i), \phi_{0,i}^N, u_{k,i}; \mathbf{q}_i),$$

which denotes the discretization of the output of the model.

We also choose a sufficiently large value for M , the number of nodes, $\{\mathbf{q}_j\}_{j=1}^M$. Therefore, the optimization problem is reduced to a standard constrained estimation problem over Euclidean M -space, in which we determine the values of the weights p_j at each node \mathbf{q}_j with the constraints that they all be nonnegative and sum to one. By Eq (3.7) it follows that

$$\begin{aligned} \hat{P}_{\mathbf{n},m,M}^N &= \arg \max_{P \in \mathcal{P}_M(Q)} \mathcal{L}_{\mathbf{n},m}^N(P; \hat{\mathbf{y}}) \\ &= \arg \max_{P \in \mathcal{P}_M(Q)} \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - y_{k,i}^N(\mathbf{q}_i)) dP(\mathbf{q}_i) \\ &= \arg \max_{P \in \mathcal{P}_M(Q)} \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N(\mathbf{q}_i), \phi_{0,i}^N, u_{k,i}; \mathbf{q}_i)) dP(\mathbf{q}_i) \\ &= \arg \max_{\tilde{\mathbf{p}} \in \mathbb{R}^M} \prod_{i=1}^m \sum_{j=1}^M \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N, \phi_{0,i}^N, u_{k,i}; \mathbf{q}_j)) p_j \\ &= \arg \max_{\tilde{\mathbf{p}} \in \mathbb{R}^M} \prod_{i=1}^m \sum_{j=1}^M p_j \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N, \phi_{0,i}^N, u_{k,i}; \mathbf{q}_j)), \end{aligned}$$

where $\tilde{\mathbf{p}} = (p_1, \dots, p_M) \in \widetilde{\mathbb{R}^M} = \{\tilde{\mathbf{p}} \mid p_j \in \mathbb{R}^+, \sum_{j=1}^M p_j = 1\}$.

We note that computing $\hat{P}_{\mathbf{n},m,M}^N$ involves high order products of very small numbers that, not unexpectedly, can cause numerical underflow. In order to mitigate this, we maximize the log-likelihood function instead and rewrite it in a form that lends itself to the use of the MATLAB optimization routine *logsumexp* as follows:

$$\begin{aligned} \hat{P}_{\mathbf{n},m,M}^N &= \arg \max_{\tilde{\mathbf{p}} \in \mathbb{R}^M} \log \left(\prod_{i=1}^m \sum_{j=1}^M p_j \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N, \phi_{0,i}^N, u_{k,i}; \mathbf{q}_j)) \right) \quad (5.2) \\ &= \arg \max_{\tilde{\mathbf{p}} \in \mathbb{R}^M} \sum_{i=1}^m \log \left(\sum_{j=1}^M p_j \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N, \phi_{0,i}^N, u_{k,i}; \mathbf{q}_j)) \right) \\ &= \arg \max_{\tilde{\mathbf{p}} \in \mathbb{R}^M} \sum_{i=1}^m \log \left(\sum_{j=1}^M \exp \left(\log \left(p_j \prod_{k=1}^{n_i} \varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N, \phi_{0,i}^N, u_{k,i}; \mathbf{q}_j)) \right) \right) \right) \\ &= \arg \max_{\tilde{\mathbf{p}} \in \mathbb{R}^M} \sum_{i=1}^m \log \left(\sum_{j=1}^M \exp \left(\log(p_j) + \sum_{k=1}^{n_i} \log(\varphi(\hat{y}_{k,i} - h_k(x_{k,i}^N, \phi_{0,i}^N, u_{k,i}; \mathbf{q}_j))) \right) \right). \end{aligned}$$

6. Abstract parabolic systems

In order to apply our estimation theory to Eqs (1.1)–(1.5), our model for the transdermal transport of ethanol given in Section 1, we reformulate it as an abstract parabolic system. We briefly describe what an abstract parabolic system is, its properties and its finite dimensional approximation, and then we show how assumptions **A1–A3** are satisfied for such a system.

Let H and V be Hilbert spaces with V densely and continuously embedded in H . Pivoting on H , it follows that H is therefore densely and continuously embedded in the dual of V , V^* . This is known as a Gelfand triple and is generally written as $V \hookrightarrow H \hookrightarrow V^*$ [37]. An abstract parabolic system is a dynamical system of the following form

$$\begin{aligned} \langle \dot{x}, \psi \rangle_{V^*, V} + a(\mathbf{q}; x, \psi) &= \langle \mathbf{B}(\mathbf{q})u, \psi \rangle_{V^*, V}, \quad \psi \in V, \\ x(0) &= x_0, \\ y(t) &= \mathbf{C}(\mathbf{q})x(t), \end{aligned} \quad (6.1)$$

where $\langle \cdot, \cdot \rangle_{V^*, V}$ denotes the duality pairing between V^* and V , \mathcal{Q} is as defined in Section 1, and for each $\mathbf{q} \in \mathcal{Q}$, $a(\mathbf{q}; \cdot, \cdot) : V \times V \rightarrow \mathbb{C}$ is a sesquilinear form satisfying the following three assumptions:

B1. (Boundedness) There exists a constant α_0 such that for all $\psi_1, \psi_2 \in V$, we have

$$|a(\mathbf{q}; \psi_1, \psi_2)| \leq \alpha_0 \|\psi_1\|_V \|\psi_2\|_V.$$

B2. (Coercivity) There exists $\lambda_0 \in \mathbb{R}$ and $\mu_0 > 0$ such that for all $\psi \in V$, we have

$$a(\mathbf{q}; \psi, \psi) + \lambda_0 |\psi|_H^2 \geq \mu_0 \|\psi\|_V^2.$$

B3. (Continuity) For all $\psi_1, \psi_2 \in V$ and $\mathbf{q}, \tilde{\mathbf{q}} \in \mathcal{Q}$, we have

$$|a(\mathbf{q}, \psi_1, \psi_2) - a(\tilde{\mathbf{q}}, \psi_1, \psi_2)| \leq d(\mathbf{q}, \tilde{\mathbf{q}}) \|\psi_1\|_V \|\psi_2\|_V.$$

In these assumptions, $\|\cdot\|_V$ and $|\cdot|_H$ denote the norm on the spaces V and H , respectively. Further, in (6.1), $\mathbf{B}(\mathbf{q}) : \mathbb{R}^V \rightarrow V^*$, and $\mathbf{C}(\mathbf{q}) : V \rightarrow \mathbb{R}$ are bounded linear operators with initial conditions $x_0 \in H$, input $u \in L^2([0, T], \mathbb{R}^V)$, and output $y \in L^2([0, T], \mathbb{R})$.

It can be shown that the system in (6.1) has a unique solution in

$$\{\psi \mid \psi \in L^2([0, T], V), \dot{\psi} \in L^2([0, T], V^*)\} \subset C([0, T], H)$$

using standard variational arguments (such as in [38]). However, we use a linear semigroup approach to convert the system in (6.1) into a discrete-time state space model, and then use arguments from linear semigroup theory [6, 39] to argue convergence of finite dimensional Galerkin-based approximations and conclude that assumptions **A1–A3** are satisfied.

Assumptions **B1** and **B2** yield that the form $a(\mathbf{q}; \cdot, \cdot)$ defines a bounded linear operator $\mathbf{A}(\mathbf{q}) : V \rightarrow V^*$ given by

$$\langle \mathbf{A}(\mathbf{q})\psi_1, \psi_2 \rangle_{V^*, V} = -a(\mathbf{q}; \psi_1, \psi_2),$$

for $\psi_1, \psi_2 \in V$. If we restrict the operator $\mathbf{A}(\mathbf{q})$ to the subspace

$$\text{Dom}(\mathbf{A}(\mathbf{q})) = \{\psi \in V \mid \mathbf{A}(\mathbf{q})\psi \in H\},$$

it becomes the infinitesimal generator of a holomorphic or analytic semigroup $\{e^{\mathbf{A}(\mathbf{q})t} \mid t \geq 0\}$ of bounded linear operators on H . The operator $\mathbf{A}(\mathbf{q})$ is referred to as being regularly dissipative [5, 6, 37]. Moreover, this semigroup can also be extended and restricted to be a holomorphic semigroup on V^* and V , respectively [5, 37].

The system in (6.1) can now be written in state space form with time invariant operators $\mathbf{A}(\mathbf{q})$, $\mathbf{B}(\mathbf{q})$ and $\mathbf{C}(\mathbf{q})$ as

$$\begin{aligned} \dot{x}(t) &= \mathbf{A}(\mathbf{q})x(t) + \mathbf{B}(\mathbf{q})u(t), \\ x(0) &= x_0, \\ y(t) &= \mathbf{C}(\mathbf{q})x(t). \end{aligned} \quad (6.2)$$

The operator form of the variation of constants formula then yields what is known as a mild solution of (6.2) given by

$$\begin{aligned} x(t; \mathbf{q}) &= e^{\mathbf{A}(\mathbf{q})t} x_0 + \int_0^t e^{\mathbf{A}(\mathbf{q})(t-s)} \mathbf{B}(\mathbf{q})u(s) ds, \quad t \geq 0, \\ y(t; \mathbf{q}) &= \mathbf{C}(\mathbf{q})x(t; \mathbf{q}). \end{aligned} \quad (6.3)$$

To obtain the corresponding discrete or sampled time form of the system given in (6.2) or (6.3), let $\tau > 0$ be the length of the sampling interval and consider strictly zero-order hold inputs of the form $u(t) = u_{k-1}$, $t \in [(k-1)\tau, k\tau)$, $k = 1, 2, \dots$. Then, let $x_k = x(k\tau)$ and $y_k = y(k\tau)$, $k = 1, 2, \dots$. By applying (6.3) on each subinterval $[(k-1)\tau, k\tau)$, $k = 1, 2, \dots$, we obtain the discrete-time dynamical system given by

$$x_k = \hat{\mathbf{A}}(\mathbf{q})x_{k-1} + \hat{\mathbf{B}}(\mathbf{q})u_{k-1}, \quad k = 1, 2, \dots, \quad (6.4)$$

$$y_k = \hat{\mathbf{C}}(\mathbf{q})x_k, \quad k = 1, 2, \dots, \quad (6.5)$$

where $x_0 \in V$, $\hat{\mathbf{A}}(\mathbf{q}) = e^{\mathbf{A}(\mathbf{q})\tau}$, $\hat{\mathbf{B}}(\mathbf{q}) = \int_0^\tau e^{\mathbf{A}(\mathbf{q})s} \mathbf{B}(\mathbf{q}) ds$ and $\hat{\mathbf{C}}(\mathbf{q}) = \mathbf{C}(\mathbf{q})$.

Using a standard Galerkin approach [40], we can approximate the discrete-time system given in (6.4)–(6.5) by a sequence of approximating finite dimensional discrete-time systems in a sequence of finite dimensional subspaces V^N of V . In order to argue convergence, we will require the following additional assumption concerning the subspaces V^N .

C1. (Approximation) For every $x \in V$, there exists $x^N \in V^N$ such that $\|x - x^N\|_V \rightarrow 0$ as $N \rightarrow \infty$.

We consider the sequence of approximating finite dimensional discrete-time systems by

$$x_k^N = \hat{\mathbf{A}}^N(\mathbf{q})x_{k-1}^N + \hat{\mathbf{B}}^N(\mathbf{q})u_{k-1}, \quad k = 1, 2, \dots,$$

$$y_k^N = \hat{\mathbf{C}}^N(\mathbf{q})x_k^N, \quad k = 1, 2, \dots,$$

where $\hat{\mathbf{A}}^N(\mathbf{q}) = e^{\mathbf{A}^N(\mathbf{q})\tau}$, $\hat{\mathbf{B}}^N(\mathbf{q}) = \int_0^\tau e^{\mathbf{A}^N(\mathbf{q})s} \mathbf{B}^N(\mathbf{q}) ds$ and $\hat{\mathbf{C}}^N(\mathbf{q})x_k = \hat{\mathbf{C}}(\mathbf{q})x_k$, where for each $\mathbf{q} \in Q$, $\mathbf{A}^N(\mathbf{q})$ is the linear operator on V^N obtained by restricting the form $a(\mathbf{q}; \cdot, \cdot)$ to $V^N \times V^N$, i.e., for $\psi_1^N, \psi_2^N \in V^N$,

$$\langle \mathbf{A}^N(\mathbf{q})\psi_1^N, \psi_2^N \rangle_{V^*, V} = -a(\mathbf{q}; \psi_1^N, \psi_2^N).$$

In addition, $\mathbf{B}^N(\mathbf{q}) = \pi^N \mathbf{B}(\mathbf{q})$, where in this definition, π^N is the natural extension of the orthogonal projection operator $\pi^N : H \rightarrow V^N$ to V^* from its dense subspace H . We also set $x_0^N = \pi^N x_0 \in V^N$.

Under the assumptions **B1–B3** and **C1** using the Trotter-Kato approximation theorem from the theory of linear semigroups of operators [39,41], we were able to conclude that $\lim_{N \rightarrow \infty} \|x_k^N - x_k\|_V = 0$ and $\lim_{N \rightarrow \infty} |y_k^N - y_k| = 0$ for each $x_0 \in V$, and uniformly in \mathbf{q} for $\mathbf{q} \in Q$ and $k \in \{1, 2, \dots, K\}$, for any fixed $K \in \mathbb{N}^+$.

We can now use the results described in the previous paragraphs to show that an abstract parabolic system satisfies assumptions **A1–A3** given in Section 5. To show that the assumption **A1** is satisfied, we need to show that for all \mathbf{n} , m , and N , the map $P \mapsto \mathcal{L}_{\mathbf{n},m}^N(P; \hat{\mathbf{y}})$ is a continuous map. It suffices to show that for any fixed \mathbf{n} , m , and N , and for any sequence of probability measures P_M , such that $\rho(P_M, P) \rightarrow 0$ in $\mathcal{P}(Q)$, we have $\mathcal{L}_{\mathbf{n},m}^N(P_M; \hat{\mathbf{y}}) \rightarrow \mathcal{L}_{\mathbf{n},m}^N(P; \hat{\mathbf{y}})$ as $M \rightarrow \infty$. Toward this end, we see that

$$\begin{aligned} \left| \mathcal{L}_{\mathbf{n},m}^N(P_M; \hat{\mathbf{y}}) - \mathcal{L}_{\mathbf{n},m}^N(P; \hat{\mathbf{y}}) \right| &= \left| \prod_{i=1}^m \mathcal{L}_i^N(P_M; \hat{\mathbf{y}}_i) - \prod_{i=1}^m \mathcal{L}_i^N(P; \hat{\mathbf{y}}_i) \right| \\ &= \left| \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}^N(\mathbf{q}_i)) dP_M(\mathbf{q}_i) \right. \\ &\quad \left. - \prod_{i=1}^m \int_Q \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}^N(\mathbf{q}_i)) dP(\mathbf{q}_i) \right| \\ &\rightarrow 0, \end{aligned}$$

by definition of the Prohorov metric. It follows that the assumption **A1** is satisfied.

Next, we show that the assumption **A2** is satisfied. We have that $\lim_{N \rightarrow \infty} \|x_{k,i}^N - x_{k,i}\|_V = 0$ and $\lim_{N \rightarrow \infty} |y_{k,i}^N - y_{k,i}| = 0$ for each $x_{0,i} \in V$, uniformly in \mathbf{q}_i for $\mathbf{q}_i \in Q$, $k = 1, \dots, n_i$, $i = 1, \dots, m$. We want to show that for any sequence of probability measures P_M , such that $\rho(P_M, P) \rightarrow 0$ in $\mathcal{P}(Q)$, and for $i = 1, \dots, m$, as $M, N \rightarrow \infty$, we have $\mathcal{L}_i^N(P_M; \hat{\mathbf{y}}_i) \rightarrow \mathcal{L}_i(P; \hat{\mathbf{y}}_i)$.

Recall that φ is assumed to be continuous. Let $\varepsilon > 0$. Choose N_0 such that for $N \geq N_0$, and for every M , $\left| \int_Q \left(\prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}^N(\mathbf{q}_i)) - \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}(\mathbf{q}_i)) \right) dP_M(\mathbf{q}_i) \right| < \varepsilon/2$. Then, we have

$$\begin{aligned} &\left| \mathcal{L}_i^N(P_M; \hat{\mathbf{y}}_i) - \mathcal{L}_i(P; \hat{\mathbf{y}}_i) \right| \\ &= \left| \int_Q \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}^N(\mathbf{q}_i)) dP_M(\mathbf{q}_i) - \int_Q \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}(\mathbf{q}_i)) dP(\mathbf{q}_i) \right| \\ &\leq \left| \int_Q \left(\prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}^N(\mathbf{q}_i)) - \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}(\mathbf{q}_i)) \right) dP_M(\mathbf{q}_i) \right| \\ &\quad + \left| \int_Q \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}(\mathbf{q}_i)) dP_M(\mathbf{q}_i) - \int_Q \prod_{k=1}^{n_i} \varphi(\hat{\mathbf{y}}_{k,i} - y_{k,i}(\mathbf{q}_i)) dP(\mathbf{q}_i) \right| \\ &< \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon, \end{aligned}$$

where the second term is less than $\varepsilon/2$ by definition of the Prohorov metric. Consequently, the assumption **A2** is satisfied.

Finally, we want to show that the assumption **A3** is satisfied. We want to show that for all $P \in \mathcal{P}(Q)$ and for $i = 1, \dots, m$, $\mathcal{L}_i^N(P_M; \hat{\mathbf{y}}_i)$ and $\mathcal{L}_i(P; \hat{\mathbf{y}}_i)$ are uniformly bounded. Recall that the parameter space Q is compact. Thus, for $\mathbf{q}_i \in Q$, and for each N , $y_{k,i}^N(\mathbf{q}_i)$ are uniformly bounded. Similarly, $y_{k,i}(\mathbf{q}_i)$ are also uniformly bounded and we also have that $|y_{k,i}^N(\mathbf{q}_i) - y_{k,i}(\mathbf{q}_i)| \rightarrow 0$ uniformly in \mathbf{q}_i for $\mathbf{q}_i \in Q$. Therefore, we can conclude that the assumption **A3** is satisfied.

7. Application to the transdermal transport of alcohol

To apply the results established in Section 6 to the system (1.1)–(1.5) in Section 1, the system must first be written in weak form, then the parameter space Q , the Hilbert spaces H and V , the sesquilinear form $a(\mathbf{q}; \cdot, \cdot)$ and the operators $\mathbf{B}(\mathbf{q})$ and $\mathbf{C}(\mathbf{q})$ must all be identified. Also, the approximating subspaces V^N must be chosen, and finally assumptions **B1–B3** and **C1** must all be shown to be satisfied.

The parameter space Q is assumed to be a compact subset of $\mathbb{R}^+ \times \mathbb{R}^+$ with any p -metric denoted by d_Q . Let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\mathbf{y} = (y_1, \dots, y_n) \in \mathbb{R}^n$, then the p -metric is defined by

$$d_p(\mathbf{x}, \mathbf{y}) = \left(\sum_{i=1}^n |x_i - y_i|^p \right)^{1/p}$$

for any $p \in [1, \infty)$. For $p = \infty$, the p -metric is defined by

$$d_p(\mathbf{x}, \mathbf{y}) = \max_{i=1, \dots, n} |x_i - y_i|.$$

Let $H = L^2(0, 1)$ and $V = H^1(0, 1)$ with their standard inner products and norms. It follows that $V^* = H^{-1}(0, 1)$ and the three spaces H , V and V^* form a Gelfand triple. To rewrite the system (1.1)–(1.5) in weak form, we multiply by a test function $\psi \in V$ and integrate by parts to obtain

$$\langle \dot{x}(t), \psi \rangle_{V^*, V} + \int_0^1 q_1 \frac{\partial x}{\partial \eta}(t, \eta) \psi'(\eta) d\eta + x(t, 0) \psi(t, 0) = q_2 u(t) \psi(1),$$

where $\langle \cdot, \cdot \rangle_{V^*, V}$ denotes the duality pairing between V^* and V . Then, for $\mathbf{q} \in Q$, $u \in \mathbb{R}$ and $\tilde{\psi}, \psi \in V$, we set

$$\begin{aligned} a(\mathbf{q}; \tilde{\psi}, \psi) &= \int_0^1 q_1 \tilde{\psi}'(\eta) \psi'(\eta) d\eta + \tilde{\psi}(0) \psi(0), \\ \langle \mathbf{B}(\mathbf{q})u, \psi \rangle_{V^*, V} &= q_2 u \psi(1), \\ \mathbf{C}(\mathbf{q})\psi &= \mathbf{C}\psi = \psi(0). \end{aligned}$$

We can establish that assumptions **B1–B3** are satisfied using arguments involving the Sobolev embedding theorem (see [42]). Also, the operators $\mathbf{B}(\mathbf{q})$ and $\mathbf{C}(\mathbf{q})$ are continuous in the uniform operator topology with respect to $\mathbf{q} \in Q$. It follows from Section 6 that

$$\begin{aligned} g_{k-1}(x_{k-1,i}(\mathbf{q}_i), u_{k-1,i}; \mathbf{q}_i) &= \hat{\mathbf{A}}(\mathbf{q}_i) x_{k-1,i}(\mathbf{q}_i) + \hat{\mathbf{B}}(\mathbf{q}_i) u_{k-1,i}, \quad k = 1, 2, \dots, n_i, \quad i = 1, \dots, m, \\ h_k(x_{k,i}(\mathbf{q}_i), \phi_{0,i}, u_{k,i}; \mathbf{q}_i) &= \hat{\mathbf{C}}(\mathbf{q}_i) x_{k,i}(\mathbf{q}_i), \quad k = 1, \dots, n_i, \quad i = 1, \dots, m, \end{aligned}$$

where $\hat{\mathbf{A}}(\mathbf{q}) = e^{\mathbf{A}(\mathbf{q})\tau}$, $\hat{\mathbf{B}}(\mathbf{q}) = \int_0^\tau e^{\mathbf{A}(\mathbf{q})s} \mathbf{B}(\mathbf{q}) ds$ and $\hat{\mathbf{C}}(\mathbf{q}) = \mathbf{C}(\mathbf{q})$, with $\tau > 0$ which is the length of the sampling interval.

Let V^N , $N = 1, 2, \dots$, be the span of the standard linear splines defined with respect to the uniform mesh $\{0, 1/N, 2/N, \dots, (N-1)/N, 1\}$ on $[0, 1]$. Then, assumption **C1** is satisfied by standard arguments for spline functions (see, for example, [43]). If for each $i = 1, 2, \dots, m$, we define $x_{k,i}^N$ and $y_{k,i}^N$ as in (6.4)–(6.5), then by the arguments at the end of Section 6, we conclude that assumptions **A1–A3** are satisfied.

In the following two subsections, 7.1 and 7.2, we present the application of our scheme to the transdermal transport of alcohol in two examples, one involving simulated data, and the other using actual human subject data collected in the Luczak laboratory at USC. For the simulated data, we want to show the convergence of the estimated distribution of the parameter vector $\mathbf{q} = (q_1, q_2)$ to the “true” distribution as the number of drinking episodes increases, as the number of nodes increases and as the level of discretization in the finite dimensional approximations increases. For the actual data we apply the LOOCV method by first estimating the distribution of the parameter vector \mathbf{q} using the training set, and then we estimated the TAC output using the estimated distribution of the parameter vector \mathbf{q} and the BrAC input of the test set.

7.1. Example 1: Estimation based on simulation data

In this example, we estimate the distribution of the parameter vector $\mathbf{q} = (q_1, q_2)$ in the system (1.1)–(1.5) by first simulating TAC data in MATLAB with the assumption that the two parameters q_1 and q_2 are i.i.d. with a Beta distribution, $q_1, q_2 \sim \text{Beta}(2, 5)$. Thus, their joint cumulative distribution function (cdf) is the product of their marginal $\text{Beta}(2, 5)$ cdfs.

From Eq (3.1), we have

$$Y_{k,i} = y_{k,i}(\mathbf{q}_i) + e_{k,i}, \quad k = 1, \dots, n_i, \quad i = 1, \dots, m,$$

where m is the number of drinking episodes and $y_{k,i}(\mathbf{q}_i)$ is the observed TAC for the i^{th} drinking episode at time step k . We let P_0 be the product of the cdfs of two independent $\text{Beta}(2, 5)$ distributions, and $e_{k,i} \sim N(0, 10^{-6})$, $k = 1, \dots, n_i$, $i = 1, \dots, m$. The reason for the small variance is because the noise accounts for errors in the measurements caused by factors that result in small perturbations to the measurements, such as humidity or temperature as external environmental factors or the level of skin moisture or greasiness among different drinking episodes. It is assumed that these types of factors are normally distributed, causing small variations to the measurements.

To approximate the PDE model for the TAC observations, we used the linear spline-based Galerkin approximation scheme described in Section 6 with N equally spaced subintervals from $[0, 1]$ (see [15–17]). We want to compute $\hat{P}_{\mathbf{n}, m, M}^N$ given by Eq (5.2), where $\mathbf{q}_j = (q_{j1}, q_{j2})$ is chosen as M uniform meshgrid coordinates on $[0, 1] \times [0, 1]$. We make the assumption that there is no alcohol in the epidermal layer of the skin at time $t = 0$, so we let $\phi_{0,i}^N = 0$. The constrained optimization problem over Euclidean M -space was solved using constrained optimization routine FMINCON (find minimum of constrained nonlinear multivariable function) from the Optimization Toolbox in MATLAB applied to the negative of the log-likelihood function.

We note that before turning to a mixed effects statistical model, we had considered an output or observation that was aggregated TAC; in this case, the appropriate underlying statistical model was

the naive pooled model (i.e., the data point for each drinking episode at a certain time is an observation of the mean behavior plus a random error). However, it is not difficult to show that when this observation is used, if the \mathbf{q}_1 and \mathbf{q}_2 are assumed to be independent, then their joint distribution is not identifiable. Consequently, when the nonlinear least squares-based constrained optimization problem is solved, the inherent ill-posedness of the inverse problem results in undesirable oscillations. To mitigate this behavior, we had to introduce an appropriately weighted regularization term in the performance index being minimized. One advantage of the mixed effects statistical model presented here is that regularization is not required.

In order to demonstrate the consistency of our estimator, we show that as m , the number of drinking episodes, increases, the estimated cdf of the parameter vector $\mathbf{q} = (q_1, q_2)$ approaches the “true” cdf, the product of two $Beta(2, 5)$ cdfs. In order to simulate realistic longitudinal TAC vectors representing data that might be collected by the TAC biosensor for an individual’s drinking episode, we used BrAC data collected in the Luczak laboratory as the input to the model and generated random samples of q_1 and q_2 , i.i.d. $Beta(2, 5)$ in MATLAB. Using the algorithm developed in the current paper, we estimated the distribution of the random parameter vector by solving the optimization problem for different cases based on the number of drinking episodes m and observed the convergence of the estimated distribution to the “true” distribution as m increases.

To quantify this, let D be the sum of the squared differences at each node between the estimated and the “true” distribution, the product of two $Beta(2, 5)$ cdfs. Let p_j and b_j be the weights at the node \mathbf{q}_j of the estimated and “true” distribution, respectively. So, we have

$$D = \sum_{j=1}^M (p_j - b_j)^2$$

representing the error in estimating the weights at each node. This is a metric being used to quantify that the difference between the estimated distribution and the “true” distribution decreases as the number of drinking episodes increases.

We fixed the number of nodes M and the level of discretization N sufficiently large. We set $M = 400$ and $N = 128$. We estimated the distribution of the parameter vector for different cases based on different numbers of drinking episodes, $m \in \{1, 3, 7, 9, 16, 42\}$, and calculated D for each case. Our results are summarized in Table 1. We observed that as the number of drinking episodes m increases, the sum of the squared differences at each node between the estimated distribution and the “true” distribution D decreases.

In Figure 2, each row of the figure contains three different views of the same plot of the estimated distribution and the “true” distribution (again, the product of two $Beta(2, 5)$ cdfs) for different numbers of drinking episodes $m = 7$, $m = 16$ and $m = 42$ in the top, middle and bottom rows, respectively, with the number of nodes set to $M = 400$ and the level of discretization to $N = 128$. We observe that as m increases, our estimated distribution gets “closer” to the “true” distribution, which agrees with the numerical results that are shown in Table 1.

Table 1. Decrease in D , the sum of the squared differences at each node between the estimated and the “true” distribution, with increasing m , the number of drinking episodes, for fixed values of the number of nodes $M = 400$ and the level of discretization $N = 128$.

m	D
1	39.0164
3	28.3091
7	8.3247
9	7.1750
16	3.5697
42	3.0337

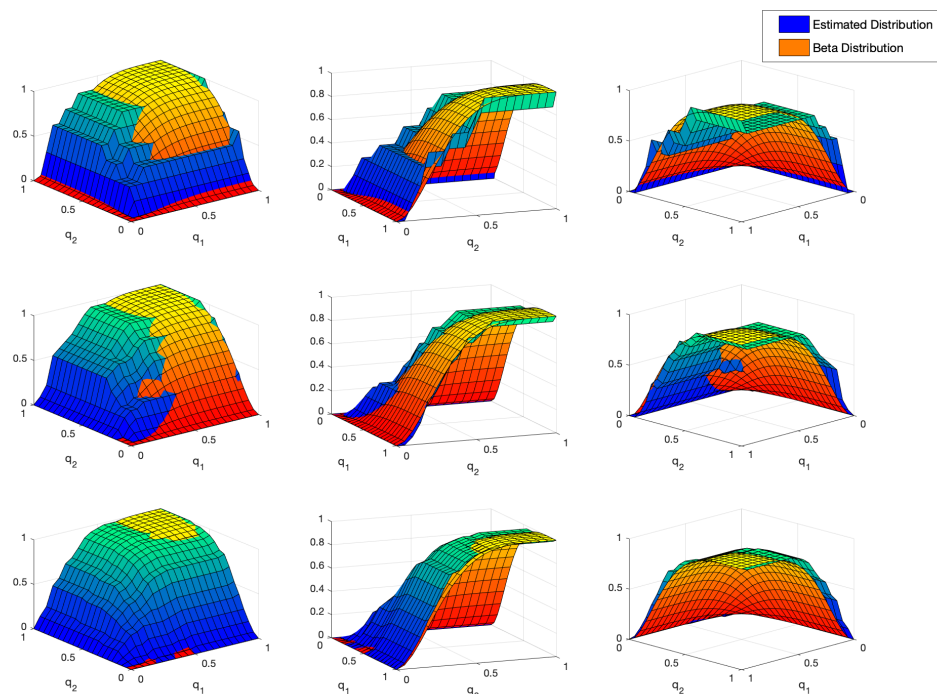


Figure 2. Each row of the figure contains three different views of the same plot of the estimated distribution and the “true” joint $Beta(2,5)$ distribution for different numbers of drinking episodes $m = 7$, $m = 16$ and $m = 42$ in the top, middle and bottom rows, respectively, for a fixed number of nodes $M = 400$ and level of discretization $N = 128$. We observe that as m increases from the top row to the bottom row, our estimated distribution gets “closer” to the “true” distribution, which agrees with the numerical results that are shown in Table 1.

Next, we show that as the number of nodes M and the level of discretization N increases, the normalized sum of squared differences at each node between the estimated and the “true” distribution decreases.

First, we fixed the level of discretization at $N = 128$ and we increased the number of nodes M . Let

$$\bar{D}_M = \frac{1}{M} \sum_{j=1}^M (p_j - b_j)^2.$$

In Table 2, we observe that for the fixed value of $N = 128$, as the number of nodes M increases, the normalized sum of the squared differences at each node between the estimated distribution and the “true” distribution \bar{D}_M decreases.

Table 2. Decrease in \bar{D}_M , the normalized sum of the squared differences at each node between the estimated distribution and the “true” distribution, with increasing M , the number of nodes, for a fixed value for the level of discretization $N = 128$.

M	\bar{D}_M
25	0.03795
100	0.03025
225	0.02974
400	0.02081

Next, we fixed the number of nodes at $M = 400$ and we increased the level of discretization N . Let

$$\bar{D}_N = \frac{1}{N} \sum_{j=1}^M (p_j - b_j)^2.$$

In Table 3, we observe that for $N = 128$ fixed, as the number of nodes M increases, the normalized sum of the squared differences at each node between the estimated distribution and the “true” distribution \bar{D}_N decreases.

Table 3. Decrease in \bar{D}_N , the normalized sum of the squared differences at each node between the estimated distribution and the “true” distribution, with increasing N , the level of discretization, with the number of nodes fixed at $M = 400$.

N	\bar{D}_N
4	1.22783
16	0.65644
64	0.18565
128	0.06504

The choice of the “true” distribution for q_1 and q_2 in the simulation case is the scatterplot of samples for a set of 18 drinking episodes, including BrAC and TAC measurements of different individuals obtained by a deterministic approach in [22]. However, the $Beta(2, 5)$ distribution was chosen strictly for the purpose of demonstration. When applying our algorithm to actual clinic or lab collected human subject data, a valuable feature of our developed methodology in using a nonparametric approach is that we do not need to make any assumptions about the family of feasible

distributions for the parameter vector unlike the parametric approach in [15–17]. In addition, the independent and identically distributed assumption was also very simplistic given that q_1 and q_2 parameters depend on the same individual and environmental conditions at the time of measurements. In the nonparametric approach, there is no need to make any assumptions about the distribution, resulting in relaxing the restrictive assumption of the distribution being independent and identically distributed. Thus, this assumption is relaxed in the flexible nonparametric approach used in the next example.

7.2. Example 2: Estimation based on actual human subject data

The two datasets used in this example were obtained by two different alcohol biosensors: SCRAM CAM[®] and WrisTAS[™]7 [44, 45]. We fixed the number of nodes at $M = 400$ and the level of discretization at $N = 128$, both sufficiently large with respect to convergence as we observed in our simulation data examples. From each dataset, we chose $m = 9$ different drinking episodes. We split the drinking episodes into a training set consisting of eight drinking episodes and a testing set consisting of one drinking episode. This way, we could apply the LOOCV method. We repeated this partitioning process nine times, each time leaving out a different drinking episode. Using the training set, we first estimated the distribution of the parameter vector $\mathbf{q} = (q_1, q_2)$. Next, we sampled 100 parameter vectors $\mathbf{q} = (q_1, q_2)$ from the estimated distribution, and using those along with the BrAC input from the testing dataset, we simulated 100 TAC longitudinal signals. From these 100 simulated TAC signals, we estimated the “true” TAC by computing the mean at each time, and we provided what we refer to as a 95% conservative error band, or simply as a 95% error band, by taking the 2.5 and 97.5 percentiles. This approach for the error band is also used for a number of statistics associated with the TAC curve that are of particular interest to researchers and clinicians working in the area of alcohol use disorder.

For the first example, we considered the dataset collected using the SCRAM alcohol biosensor. Prior to applying the LOOCV method, in order to visualize the estimated density and distribution of $\mathbf{q} = (q_1, q_2)$ and the marginal densities of q_1 and q_2 , we trained the algorithm on all nine drinking episodes. Figure 3 illustrates the four aforementioned plots. In the estimated density plot, we can see that our numerical result for this example is in agreement with the theoretical result in Lindsay and Mallet [34, 35], which states that the maximum likelihood estimator $\hat{P}_{n,m}$ can be found in the class of discrete distributions with at most m support points, i.e., $\hat{P}_{n,m} \in \mathcal{P}_M(Q)$, where $M \leq m$. In this example, since we had nine drinking episodes, the estimated density plot displays the support points among 400 nodes for $\mathbf{q} = (q_1, q_2)$.

In addition, for this sample, the sample mean of $\mathbf{q} = (q_1, q_2)$ is calculated to be

$$\bar{\mathbf{q}} = (0.6003, 1.2452),$$

the sample covariance matrix is calculated to be

$$\mathbf{S}_{\mathbf{q}} = \begin{pmatrix} 0.0706 & -0.0264 \\ -0.0264 & 0.0483 \end{pmatrix}$$

and the sample correlation is calculated to be $\rho_{\mathbf{q}} = -0.4519$. Based on this, we observe that for our training population consisting of nine drinking episodes, there is a moderate negative association

between the parameters q_1 and q_2 . Recall that the parameters q_1 and q_2 in the system (1.1)–(1.5) represent the normalized diffusivity and the normalized flux gain at the boundary between the dermal and epidermal layers, respectively. It is important to note that the negative correlation is specifically for these nine drinking episodes, and it should not be generalized as the behavior expected in a different sample. In particular, in the next example we have a moderate positive correlation. In addition, it is possible to argue heuristically why one would expect either negative or positive correlation. Significantly more data and testing would be required for a statistical analysis with sufficient power to determine if any conclusion regarding the correlation between the parameters could be drawn.

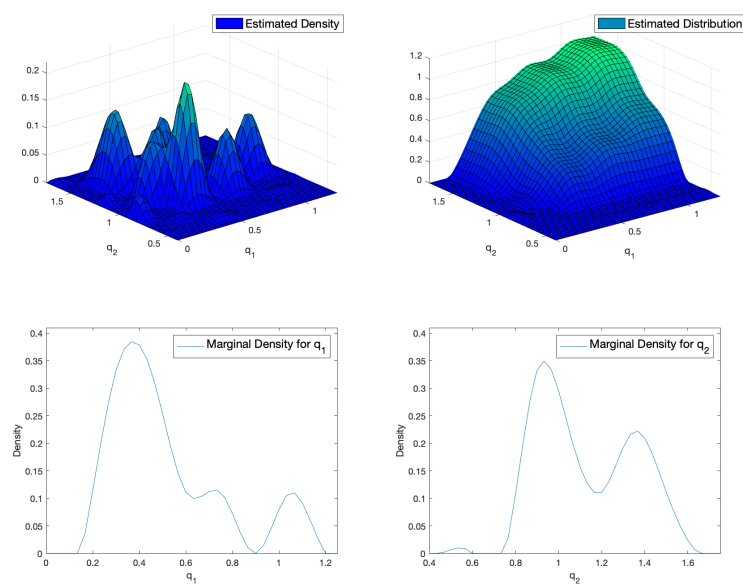


Figure 3. The estimated probability density function (pdf) (top left), the estimated cdf (top right), marginal density of q_1 (bottom left), and marginal density of q_2 (bottom right) obtained from $m = 9$ drinking episodes collected using the SCRAM alcohol biosensor, for a fixed number of nodes $M = 400$ and level of discretization $N = 128$. In the estimated density plot, we can see that our numerical result for this example is in agreement with the theoretical result in Lindsay and Mallet [34,35], which states that the maximum likelihood estimator can be found in the class of discrete distributions with at most m support points, where $M \leq m$. In this example, since we had nine drinking episodes, the estimated density plot displays the support points among 400 nodes for $\mathbf{q} = (q_1, q_2)$.

We applied the LOOCV method as explained above to the nine drinking episodes from the SCRAM biosensor. Figure 4 shows the measured TAC (i.e., measured by the SCRAM alcohol biosensor) and the estimated TAC (i.e., obtained from our algorithm) for all nine drinking episodes left out in the testing set in the partitioning process, and the conservative 95% error band for a fixed number of nodes $M = 400$ and level of discretization $N = 128$.

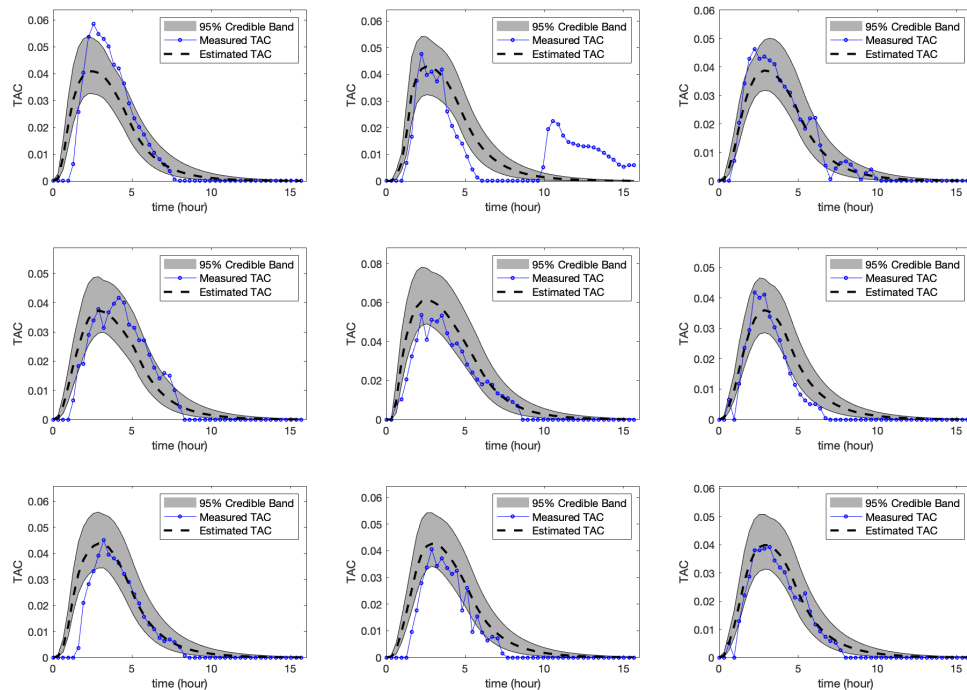


Figure 4. The measured TAC, the estimated TAC, and the conservative 95% error band for nine drinking episodes from the testing set collected using the SCRAM alcohol biosensor using the LOOCV. The figure illustrates a reasonable fit. The second peak in the middle panel of the top row was due to the individual leaving the lab on a warm day, which results in a burst of TAC. This error in the measurement was ignored by the model since this particular test episode was trained based on the remaining eight drinking episodes, which shows that TAC goes to zero as time increases.

Alcohol researchers and clinicians are particularly interested in certain statistics associated with drinking episodes: The maximum or peak value of the TAC curve, the time at which the peak value of the TAC is attained and the area under the TAC curve. The area under the curve (AUC) is a quantifying measure of exposure to the alcohol that integrates the transdermal alcohol concentration across time. Tables 4–6 display these statistics along with the measured (or actual) value obtained by the SCRAM alcohol biosensor, as well as the conservative 95% error band for the nine drinking episodes from the testing set. From these tables, we can observe that the 95% error bands do a reasonably good job of capturing the actual values of these statistics, especially for the value of the peak TAC and the area under the curve displayed in Tables 4 and 6. In Figure 4, we can see that there are minor fluctuations in the measured TAC curve. If we smooth the measured TAC curve, we will obtain better results in lowering the error between the time at which the peak value of the TAC is attained using the smoothed TAC curve and the estimated peak time compared to the results obtained in Table 5. The smoothed version results are included in Table A1 in the Appendix.

Table 4. The measured peak TAC, estimated peak TAC and the 95% error band for the nine drinking episodes from the testing set collected using the SCRAM alcohol biosensor.

Drinking episode	Measured peak TAC	Estimated peak TAC	95% Error band
1	0.0585	0.0413	(0.0327, 0.0539)
2	0.0477	0.0433	(0.0324, 0.0543)
3	0.0464	0.0391	(0.0320, 0.0501)
4	0.0417	0.0375	(0.0301, 0.0489)
5	0.0535	0.0618	(0.0497, 0.0781)
6	0.0419	0.0361	(0.0287, 0.0465)
7	0.0450	0.0441	(0.0346, 0.0558)
8	0.0405	0.0430	(0.0345, 0.0542)
9	0.0391	0.0402	(0.0313, 0.0508)

Table 5. The measured peak time, estimated peak time and the 95% error band for the nine drinking episodes from the testing set collected using the SCRAM alcohol biosensor.

Drinking episode	Measured peak time	Estimated peak time	95% Error band
1	2.5600	2.4480	(1.9200, 2.8800)
2	2.2400	2.6080	(2.2400, 2.8800)
3	2.2400	2.8928	(2.5600, 3.2000)
4	4.1600	2.9056	(2.5600, 3.2000)
5	2.2400	2.5728	(2.2400, 2.8800)
6	2.2400	2.8928	(2.5600, 3.2000)
7	3.2000	2.9696	(2.5600, 3.2000)
8	2.8800	2.9312	(2.5600, 3.2000)
9	3.2000	2.9088	(2.5600, 3.2000)

Table 6. The measured AUC, estimated AUC and 95% error band for nine drinking episodes from the testing set collected using the SCRAM alcohol biosensor.

Drinking episode	Measured AUC	Estimated AUC	95% Error band
1	0.1909	0.1784	(0.1382, 0.2229)
2	0.1876	0.1799	(0.1343, 0.2321)
3	0.1868	0.1751	(0.1364, 0.2421)
4	0.1765	0.1735	(0.1349, 0.2394)
5	0.2231	0.2963	(0.2203, 0.3911)
6	0.1151	0.1496	(0.1117, 0.1982)
7	0.1474	0.1912	(0.1444, 0.2563)
8	0.1277	0.1898	(0.1412, 0.2505)
9	0.1493	0.1750	(0.1307, 0.2319)

For the second example, we applied the LOOCV method to the nine drinking episodes from the WrisTAS^{TM7} alcohol biosensor. Figure 5 shows the measured TAC (i.e., measured by the WrisTAS^{TM7} alcohol biosensor) and the estimated TAC (i.e., obtained from our algorithm) for all nine drinking episodes left out in the testing set in the partitioning process, and the conservative 95% error band for a fixed number of nodes $M = 400$ and level of discretization $N = 128$. We can observe that we obtained similar results as the previous example which illustrates the consistency of the developed algorithm across different alcohol biosensors.

In fact, Figure 5 shows a better fit than Figure 4, and this might be due to the fact that the WrisTAS^{TM7} biosensor collected data more frequently (i.e., in a smaller time intervals) which resulted in more data points for the training purposes of the model. Another reason may be the difference between the accuracy of the biosensors, which cannot be tested using the dataset used in this paper due to several confounding factors since the two collections of nine drinking episodes in each example do not represent the same drinking episodes; that is, we do not have data collection by the two biosensors under the same conditions to eliminate the confounding factors. If the scientific question that is being addressed is the accuracy of the biosensors, then the researchers should collect data using the two biosensors simultaneously during the same drinking episode to eliminate the confounding factors.

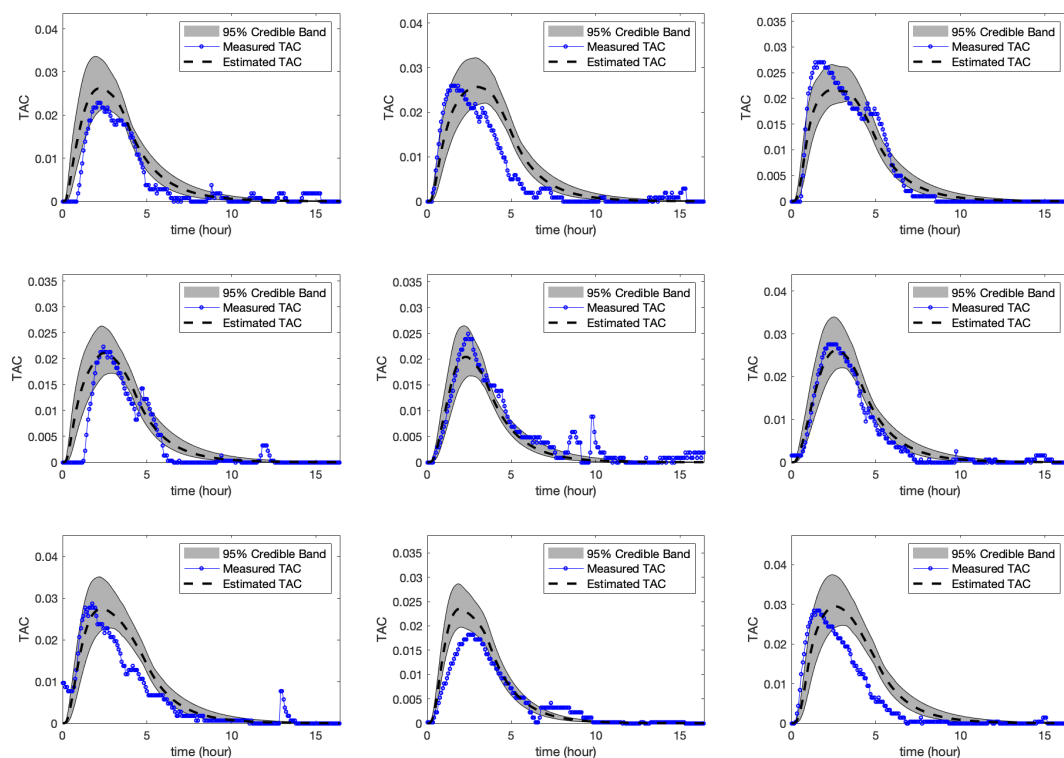


Figure 5. The measured TAC, the estimated TAC and the conservative 95% error band for nine drinking episodes from the testing set collected using the WrisTAS^{TM7} alcohol biosensor using the LOOCV. The figure illustrates a reasonable fit.

Similar to the previous example using the SCRAM biosensor, in this example, using the WrisTASTM7 biosensor, we have computed some statistics as the means of comparison between the two datasets. The sample mean of $\mathbf{q} = (q_1, q_2)$ is calculated to be

$$\bar{\mathbf{q}} = (0.5696, 0.8608),$$

the sample covariance matrix is calculated to be

$$\mathbf{S}_q = \begin{pmatrix} 0.0270 & 0.0067 \\ 0.0067 & 0.0074 \end{pmatrix},$$

and the sample correlation is calculated to be $\rho_q = 0.4771$. We observe a moderate positive association between the parameters q_1 and q_2 , while in the previous example we observed a moderate negative association between the parameters. The statistics obtained in the second example (i.e., using the WrisTASTM7 biosensor) may not be comparable to the statistics obtained in the first example (i.e., using the SCRAM biosensor) due to the fact that the two sets of nine drinking episodes are not identical; in other words, we lack data gathered by the two biosensors in identical conditions to eliminate the potential confounding factors. In addition, the sample size of nine drinking episodes is a small sample size to make any general hypothesis regarding the correlation between the two parameters q_1 and q_2 .

Tables 7–9 display these statistics along with the measured (or actual) value obtained by the WrisTASTM7 alcohol biosensor, as well as the conservative 95% error band for the nine drinking episodes from the testing set. From these tables, we can observe that the 95% error bands do a reasonably good job of capturing the actual values of these statistics, especially for the value of the peak TAC and the area under the curve displayed in Tables 7 and 9. In Figure 5, we can see that there are minor fluctuations in the measured TAC curve. Similar to the previous example, if we smooth the measured TAC curve, we will obtain better results in lowering the error between the time at which the peak value of the TAC is attained using the smoothed TAC curve and the estimated peak time compared to the results obtained in Table 8. The smoothed version results are included in Table A2 in the Appendix.

Table 7. The measured peak TAC, estimated peak TAC and the 95% error band for the nine drinking episodes from the testing set collected using the WrisTASTM7 alcohol biosensor.

Drinking episode	Measured peak TAC	Estimated peak TAC	95% Error band
1	0.0228	0.0265	(0.0215, 0.0336)
2	0.0259	0.0258	(0.0221, 0.0319)
3	0.0270	0.0218	(0.0195, 0.0266)
4	0.0222	0.0212	(0.0173, 0.0263)
5	0.0249	0.0205	(0.0168, 0.0264)
6	0.0276	0.0263	(0.0221, 0.0337)
7	0.0287	0.0276	(0.0228, 0.0348)
8	0.0182	0.0236	(0.0197, 0.0286)
9	0.0284	0.0296	(0.0247, 0.0371)

Table 8. The measured peak time, estimated peak time and the 95% error band for the nine drinking episodes from the testing set collected using the WrisTASTM7 alcohol biosensor.

Drinking episode	Measured peak time	Estimated peak time	95% Error band
1	2.1667	2.2758	(1.9167, 2.6667)
2	1.6000	3.0125	(2.8333, 3.5000)
3	1.6944	2.5542	(2.0833, 3.3333)
4	2.4167	2.5083	(2.3333, 2.9167)
5	2.4167	2.3117	(2.1229, 2.5833)
6	2.4167	2.6917	(2.5000, 3.0000)
7	1.7500	2.3358	(2.0833, 2.8333)
8	2.5833	1.9383	(1.8333, 2.0833)
9	1.5208	2.6233	(2.3333, 3.0833)

Table 9. The measured AUC, estimated AUC and 95% error band for nine drinking episodes from the testing set collected using the WrisTASTM7 alcohol biosensor.

Drinking episode	Measured AUC	Estimated AUC	95% Error band
1	0.0755	0.1054	(0.0912, 0.1229)
2	0.0937	0.1205	(0.1057, 0.1438)
3	0.1091	0.1022	(0.0927, 0.1240)
4	0.0653	0.0876	(0.0759, 0.1023)
5	0.0902	0.0730	(0.0647, 0.0871)
6	0.0942	0.1044	(0.0917, 0.1248)
7	0.1036	0.1213	(0.1063, 0.1448)
8	0.0716	0.0876	(0.0760, 0.1038)
9	0.0942	0.1298	(0.1138, 0.1549)

8. Discussion and concluding remarks

In this paper, we considered the nonparametric fitting of a population model for the transdermal transport of alcohol based on a maximum likelihood approach applied to a mixed effects statistical model. In estimating a population model, we were actually estimating the distribution of the model parameters and, consequently, the maximum likelihood estimation (MLE) problem was formulated as an optimization problem over a space of feasible probability measures endowed with the weak topology induced by the Prohorov metric. A notable advantage of our developed methodology is that by considering a nonparametric approach, there is no need to make any assumptions about the distribution being estimated.

By using a first principles physics-based model in the form of a one-dimensional diffusion equation, we were able to capture the essential features of transdermal transport while keeping the dimension of the parameter space low. In this way, we were able to avoid having to introduce regularization so as to mitigate ill-posedness and overfitting. On the other hand, the fact that the model was infinite

dimensional being based on a partial differential equation, computing the MLE necessitated finite dimensional approximation.

We were able to first theoretically demonstrate the existence and then the consistency of our MLE using a decades old result from the literature. The consistency result is with respect to the uncertainty across subjects. It is likely that the consistency results proved in [22] and [21], in the context of a naive pooled statistical model based on a nonlinear least squares estimator, for problems either the same as, or very similar to the one we consider here, would apply for the uncertainty within each subject (i.e., as the resolution of the data with respect to time increases). At present, this is just a hypothesis and a possible avenue for future research; as of yet, we have not carefully examined this possibility.

In addition, we were able to use linear semigroup theory, in particular the Trotter-Kato theorem, the properties of the weak topology and the Prohorov metric on the space of feasible probability measures to establish a convergence result with respect to the MLE for the finite dimensional approximating estimation problems and the MLE for the estimation problem posed in terms of the original underlying infinite dimensional model.

We were able to demonstrate the efficacy of our theoretical results numerically first on an example involving simulated data, and then on one involving actual human subject data from an National Institute of Health (NIH) funded study. We used our scheme to obtain the joint density and distribution of the parameters, as well as estimates and conservative 95% error bands for the TAC signal and a number of TAC related statistics of particular interest to researchers and clinicians who work in the area of alcohol use disorder.

A similar methodology has been developed in [46] using a method called the nonparametric adaptive grid (NPAG) algorithm. In their approach, the grid of support points is iteratively changed in the algorithm, which may add to the computational cost. Also as mentioned in [46], condensing the grid around the nodes with the highest probability may result in estimating only a local maximum likelihood, which may not capture the “true” distribution to be estimated in the case of a multimodal distribution. One additional difference is that their method was tested on simulated data using a normal distribution for the parameters and it has not been tested on a human subject data. Instead of condensing the grid, we use a fixed grid and directly maximized the log-likelihood function using a mixed-effects model which was tested on both simulated data and on human subject data.

In addition to consistency with respect to the intrinsic uncertainty, other extensions we are currently looking at include the development of a general framework for estimating random parameters in general finite or infinite dimensional, continuous or discrete-time dynamical systems (e.g., ODEs (ordinary differential equations), PDEs, FDEs (functional differential equations), DEs (difference equations), and more) that would potentially subsume the results presented here as well as in [22] and [21].

Finally, although correlated, TAC and BrAC are not quantitatively equivalent. BrAC is a good indicator of BAC, which is an informative quantitative measure of the level of alcohol in the body at the moment that is affecting judgment/decision-making, motor coordination, and cognition. On the other hand, TAC is temporally delayed from BrAC since it must pass through the skin prior to being measured, and TAC levels vary across individuals, devices, and within and between drinking episodes, unlike BrAC, which is relatively consistent across people and conditions. This makes TAC difficult to interpret in relation to alcohol levels in the blood at a given time, limiting its usefulness as a quantitative measure of alcohol in the body. Thus, having a way to estimate BrAC from TAC would increase its

utility for researchers, clinicians, and individuals who want quantitatively and temporally meaningful information about alcohol levels in the body.

Since the actual motivation for this investigation is the development of schemes for converting biosensor measured TAC into BAC/BrAC, the next step would be to examine how well population models, estimated using the approach we have presented here, perform when used as part of a scheme that deconvolves an estimate for BAC/BrAC from the TAC signal. In particular, we are interested in comparing it to the schemes, used for this same purpose, developed and implemented in [11,12,28–30]. In addition, we are also interested in examining how our uncertainty quantification scheme for the TAC to BAC/BrAC conversion problem performs when compared to the non-physics-based, machine learning inspired schemes developed in [18–20,33].

Use of AI tools declaration

The authors declare they have not used Artificial Intelligence (AI) tools in the creation of this article.

Acknowledgments

We thank the Luczak laboratory students and staff members, particularly Emily Saldich, for their assistance with data collection and management for the SCRAM biosensor. We also thank Dr. Tamara Wall for providing the data for the WrisTASTM7 biosensor.

This study was funded in part by the National Institute on Alcohol Abuse and Alcoholism (Grant Numbers: R21AA017711 and R01AA026368, S.E.L. and I.G.R.) and by support from the USC Women in Science and Engineering (WiSE) program (L.A.).

Conflict of interest

The authors declare no conflicts of interest.

References

1. D. A. Labianca, The chemical basis of the Breathalyzer: A critical analysis, *J. Chem. Educ.*, **67** (1990), 259–261. <https://doi.org/10.1021/ed067p259>
2. J. T. Sakai, S. K. Mikulich-Gilbertson, R. J. Long, T. J. Crowley, Validity of transdermal alcohol monitoring: fixed and self-regulated dosing, *Alcohol.: Clin. Exp. Res.*, **30** (2006), 26–33. <https://doi.org/10.1111/j.1530-0277.2006.00004.x>
3. R. M. Swift, Transdermal alcohol measurement for estimation of blood alcohol concentration, *Alcohol.: Clin. Exp. Res.*, **24** (2000), 422–423.
4. P. R. Marques, A. S. McKnight, Field and laboratory alcohol detection with 2 types of transdermal devices, *Alcohol.: Clin. Exp. Res.*, **33** (2009), 703–711. <https://doi.org/10.1111/j.1530-0277.2008.00887.x>
5. H. T. Banks, K. Ito, Approximation in LQR problems for infinite dimensional systems with unbounded input operators, *J. Math. Syst. Estim. Control*, **7** (1997), 1–34.

6. H. T. Banks, K. Kunisch, *Estimation Techniques for Distributed Parameter Systems*, Birkhauser, Boston, (1989).
7. Z. Dai, I. G. Rosen, C. Wang, N. P. Barnett, S. E. Luczak, Using drinking data and pharmacokinetic modeling to calibrate transport model and blind deconvolution-based data analysis software for transdermal alcohol biosensors, *Math. Biosci. Eng.*, **13** (2016), 911–934. <https://doi.org/10.3934/mbe.2016023>
8. M. A. Dumett, I. G. Rosen, J. Sabat, A. Shaman, L. Tempelman, C. Wang, et al., Deconvolving an estimate of breath measured blood alcohol concentration from biosensor collected transdermal ethanol data, *Appl. Math. Comput.*, **196** (2008), 724–743.
9. I. G. Rosen, S. E. Luczak, J. Weiss, Blind deconvolution for distributed parameter systems with unbounded input and output and determining blood alcohol concentration from transdermal biosensor data, *Appl. Math. Comput.*, **231** (2014), 357–376.
10. W. F. Smith, J. Hashemi, F. Presuel-Moreno, *Foundations of Materials Science and Engineering*, 3rd edition, McGraw-Hill, New York, (2004).
11. M. Allayioti, C. Oszkinat, E. Saldich, L. Goldstein, S. E. Luczak, C. Wang, et al., Parametric and non-parametric estimation of a random diffusion equation-based population model for deconvolving blood/breath alcohol concentration from transdermal alcohol biosensor data with uncertainty quantification, in *American Control Conference (ACC)*, (2023). <https://doi.org/10.23919/ACC55779.2023.10156287>
12. K. Hawekotte, S. E. Luczak, I. G. Rosen, A Bayesian approach to quantifying uncertainty in transport model parameters for, and breath alcohol concentration deconvolved from, biosensor measured transdermal alcohol level, *Math. Biosci. Eng.*, **18** (2021), 6739–6770.
13. H. Liu, L. Goldstein, S. E. Luczak, I. G. Rosen, Confidence bands for evolution systems described by parameter-dependent analytic semigroups, in *SIAM Conference on Control and its Applications*, (2023). <https://doi.org/10.1137/1.9781611977745.17>
14. H. Liu, L. Goldstein, S. E. Luczak, I. G. Rosen, Delta-method induced confidence bands for a parameter-dependent evolution system with application to transdermal alcohol concentration monitoring, in *Conference on Decision and Control*, (2023).
15. M. Sirlanci, S. E. Luczak, C. E. Fairbairn, D. Kang, R. Pan, X. Yu, et al., Estimating the distribution of random parameters in a diffusion equation forward model for a transdermal alcohol biosensor, *Automatica*, **106** (2019), 101–109. <https://doi.org/10.1016/j.automatica.2019.04.026>
16. M. Sirlanci, S. E. Luczak, I. G. Rosen, Approximation and convergence in the estimation of random parameters in linear holomorphic semigroups generated by regularly dissipative operators, in *American Control Conference (ACC)*, (2017), 3171–3176. <https://doi.org/10.23919/ACC.2017.7963435>
17. M. Sirlanci, S. E. Luczak, I. G. Rosen, Estimation of the distribution of random parameters in discrete time abstract parabolic systems with unbounded input and output: approximation and convergence, *Commun. Appl. Anal.*, **23** (2019), 287–329. <https://doi.org/10.12732/caa.v23i2.4>

18. C. Oszkinat, T. Shao, C. Wang, I. G. Rosen, A. D. Rosen, E. Saldich, et al., Estimation and uncertainty quantification via forward and inverse filtering for a covariate-dependent, physics-informed, hidden Markov model, *Inverse Probl.*, **38** (2022). <https://doi.org/10.1088/1361-6420/ac5ac7>
19. C. Oszkinat, S. E. Luczak, I. G. Rosen, Uncertainty quantification in estimating blood alcohol concentration from transdermal alcohol level with physics-informed neural networks, *IEEE Trans. Neural Networks Learn. Syst.*, **34** (2023), 8094–8101. <https://doi.org/10.1109/tnnls.2022.3140726>
20. C. Oszkinat, S. E. Luczak, I. G. Rosen, An abstract parabolic system-based physics-informed long short-term memory network for estimating breath alcohol concentration from transdermal alcohol biosensor data, *Neural Comput. Appl.*, **34** (2022), 1–19. <https://doi.org/10.1007/s00521-022-07505-w>
21. H. T. Banks, W. C. Thompson, *Least Squares Estimation of Probability Measures in the Prohorov Metric Framework*, Technical report, (2012).
22. H. T. Banks, K. B. Flores, I. G. Rosen, E. M. Rutter, M. Sirlanci, W. C. Thompson, The Prohorov metric framework and aggregate data inverse problems for random PDEs, *Commun. Appl. Anal.*, **22** (2018), 415–446.
23. M. Davidian, D. Giltinan, *Nonlinear Models for Repeated Measurement Data*, Chapman and Hall, New York, (1995).
24. M. Davidian, D. M. Giltinan, Nonlinear models for repeated measurement data: An overview and update, *Agric. Biol. Environ. Stat.*, **8** (2003), 387–419. <https://doi.org/10.1198/1085711032697>
25. E. Demidenko, *Mixed Models, Theory and Applications*, 2nd edition, John Wiley and Sons, Hoboken, (2013).
26. M. Lovern, M. Sargentini-Maier, C. Otoul, J. Watelet, Population pharmacokinetic and pharmacodynamic analysis in allergic diseases, *Drug Metab. Rev.*, **41** (2009), 475–485. <https://doi.org/10.1080/10837450902891543>
27. R. Tatarinova, M. Neely, J. Bartroff, M. van Guilder, W. Yamada, D. Bayard, et al., Two general methods for population pharmacokinetic modeling: non-parametric adaptive grid and non-parametric Bayesian, *J. Pharmacokinet. Pharmacodyn.*, **40** (2013), 189–199. <https://doi.org/10.1007/s10928-013-9302-8>
28. J. Li, S. E. Luczak, I. G. Rosen, Comparing a distributed parameter model-based system identification technique with more conventional methods for inverse problems, *J. Inverse Ill-Posed Probl.*, **27** (2019), 703–717. <https://doi.org/10.1515/jiip-2018-0006>
29. M. Sirlanci, I. G. Rosen, S. E. Luczak, C. E. Fairbairn, K. Bresin, D. Kang, 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 Probl.*, **34** (2018), 125006. <https://doi.org/10.1088/1361-6420/aae791>
30. M. Yao, S. E. Luczak, I. G. Rosen, Tracking and blind deconvolution of blood alcohol concentration from transdermal alcohol biosensor data: A population model-based LQG approach in Hilbert space, *Automatica*, **147** (2023). <https://doi.org/10.1016/j.automatica.2022.110699>

31. D. M. Dougherty, N. E. Charles, A. Acheson, S. John, R. M. Furr, N. Hill-Kapturczak, Comparing the detection of transdermal and breath alcohol concentrations during periods of alcohol consumption ranging from moderate drinking to binge drinking, *Exp. Clin. Psychopharmacol.*, **20** (2012), 373–81. <https://doi.org/10.1037/a0029021>
32. D. M. Dougherty, T. E. Karns, J. Mullen, Y. Liang, S. L. Lake, J. D. Roache, et al., Transdermal alcohol concentration data collected during a contingency management program to reduce at-risk drinking, *Drug Alcohol Depend.*, **148** (2015), 77–84. <https://doi.org/10.1016/j.drugalcdep.2014.12.021>
33. C. E. Fairbairn, D. Kang, N. Bosch, Using machine learning for real-time BAC estimation from a new-generation transdermal biosensor in the laboratory, *Drug Alcohol Depend.*, **216** (2021), 108205. <https://doi.org/10.1016/j.drugalcdep.2020.108205>
34. B. Lindsay, The geometry of mixture likelihoods: a general theory, *Ann. Stat.*, **11** (1983), 86–94. <https://doi.org/10.1214/aos/1176346059>
35. A. Mallet, A maximum likelihood estimation method for random coefficient regression models, *Biometrika*, **73** (1986), 645–656. <https://doi.org/10.2307/2336529>
36. J. Kiefer, J. Wolfowitz, Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters, *Ann. Math. Stat.*, **27** (1956), 887–906. <https://doi.org/10.1214/aoms/1177728066>
37. H. Tanabe, *Equations of Evolution (Monographs and Studies in Mathematics)*, Pitman Publishing, (1979).
38. J. L. Lions, *Optimal Control of Systems Governed by Partial Differential Equations*, Springer Berlin, Heidelberg, (1971).
39. A. Pazy, *Semigroups of Linear Operators and Applications to Partial Differential Equations*, Springer, New York, (1983).
40. H. T. Banks, K. Kunisch, The linear regulator problem for parabolic systems, *SIAM J. Control Optim.*, **22** (1984), 684–698. <https://doi.org/10.1137/0322043>
41. H. T. Banks, K. Ito, *A Unified Framework for Approximation in Inverse Problems for Distributed Parameter Systems*, NASA. Hampton, VA. Technical Reports NASA-CR-181621, (1988).
42. R. A. Adams, J. J. F. Fournier, *Sobolev Spaces*, Elsevier, (2003).
43. M. H. Schultz, *Spline Analysis*, Prentice-Hall, (1973).
44. S. E. Luczak, I. G. Rosen, T. L. Wall, Development of a real-time repeated-measures assessment protocol to capture change over the course of drinking episodes, *Alcohol Alcohol.*, **50** (2015), 1–8. <https://doi.org/10.1093/alcalc/agu100>
45. E. B. Saldich, C. Wang, I. G. Rosen, L. Goldstein, J. Bartroff, R. M. Swift, et al., Obtaining high-resolution multi-biosensor data for modeling transdermal alcohol concentration data, *Alcohol.: Clin. Exp. Res.*, **44** (2020). <https://doi.org/10.1111/acer.14358>
46. A. Kryshchenko, M. Sirlanci, B. Vader, Nonparametric estimation of blood alcohol concentration from transdermal alcohol measurements using alcohol biosensor devices, *Adv. Data Sci. Adapt. Anal.*, **26** (2021), 329–360. https://doi.org/10.1007/978-3-030-79891-8_13

Appendix

As mentioned in Section 7.2, smoothing the measured TAC curve results in lowering the error between the time at which the peak value of the TAC is attained using the smoothed TAC curve and the estimated peak time. Here, we include the results using the smoothed version of the TAC curve. The following Tables A1 and A2 correspond to Tables 5 and 8, respectively, with the smoothed TAC curve used instead of the measured TAC curve.

Table A1. The peak time using the smoothed TAC curve, estimated peak time, and the 95% error band for the nine drinking episodes from the testing set collected using the SCRAM alcohol biosensor.

Drinking episode	Smoothed peak time	Estimated peak time	95% Error band
1	3.2000	2.4480	(1.9200, 2.8800)
2	2.8800	2.6080	(2.2400, 2.8800)
3	2.8800	2.8928	(2.5600, 3.2000)
4	3.8400	2.9056	(2.5600, 3.2000)
5	3.2000	2.5728	(2.2400, 2.8800)
6	2.8800	2.8928	(2.5600, 3.2000)
7	3.5200	2.9696	(2.5600, 3.2000)
8	3.5200	2.9312	(2.5600, 3.2000)
9	3.2000	2.9088	(2.5600, 3.2000)

Table A2. The peak time using the smoothed TAC curve, estimated peak time, and the 95% error band for the nine drinking episodes from the testing set collected using the WrisTASTM7 alcohol biosensor.

Drinking episode	Smoothed peak time	Estimated peak time	95% Error band
1	2.6667	2.2758	(1.9167, 2.6667)
2	2.1667	3.0125	(2.8333, 3.5000)
3	2.3333	2.5542	(2.0833, 3.3333)
4	2.9167	2.5083	(2.3333, 2.9167)
5	2.6667	2.3117	(2.1229, 2.5833)
6	2.6667	2.6917	(2.5000, 3.0000)
7	2.0833	2.3358	(2.0833, 2.8333)
8	2.6667	1.9383	(1.8333, 2.0833)
9	2.0833	2.6233	(2.3333, 3.0833)

To smooth the TAC curve, we used cubic splines and we measured the discrepancy between the smoothed peak time and the estimated peak time using the mean squared error. For the SCRAM biosensor, the mean squared error decreased from 0.314 (prior to smoothing) to 0.293 (after smoothing). For the WrisTASTM7 biosensor, the mean squared error decreased from 0.535 (prior to smoothing) to 0.233 (after smoothing).



AIMS Press

© 2023 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0>)