

Fast Algorithms for Nonparametric Population Modeling of Large Data Sets

Gianluigi Pillonetto ^a, Giuseppe De Nicolao ^b, Marco Chierici ^a, Claudio Cobelli ^a

^a*Dipartimento di Ingegneria dell'Informazione, University of Padova, Italy*

^b*Dipartimento di Informatica e Sistemistica, University of Pavia, Italy*

Abstract

Population models are widely applied in biomedical data analysis since they characterize both the average and individual responses of a population of subjects. In absence of a reliable mechanistic model, one can resort to the Bayesian nonparametric approach that models the individual curves as Gaussian processes. This paper develops an efficient computational scheme for estimating the average and individual curves from large data sets collected in standardized experiments, i.e. with a fixed sampling schedule. It is shown that the overall scheme exhibits a “client-server” architecture. The server is in charge of handling and processing the collective data base of past experiments. The clients ask the server for the information needed to reconstruct the individual curve in a single new experiment. This architecture allows the clients to take advantage of the overall data set without violating possible privacy and confidentiality constraints and with negligible computational effort.

Key words: Nonparametric identification; Bayesian estimation; glucose metabolism; Gaussian processes; estimation theory

1 Introduction

One of the most interesting identification problems arising in biomedical data analysis is the characterization of a population of subjects. Classical examples are found in pharmacokinetics (PK) and pharmacodynamics (PD), where multiple subjects are sampled in order to obtain both the average and individual response to the administered drug. If a sufficiently large number of samples are collected in each individual, it is possible to identify a distinct model for each subject. The typical response of the population could then be obtained from the distribution of the individual models. However, the specific nature of biomedical experiments often poses technological, cost or ethical constraints that permits to collect only few data in each single subject. When the separate identification of individual models is not viable, an effective solution is provided by so-called population modeling approaches [1], [2], [3]. Such methods process all the data simultaneously in order to achieve both the typical and individual models. Although originated in the PK/PD field, population modeling is becoming more and more popular

also in other scenarios as metabolic systems, medical imaging and even genomics [4], [5], [6].

The standard population model is a continuous-time dynamical system containing a finite number of unknown parameters, typically a compartmental model [7]. This leads to a nonlinear-in-parameter identification problem that can be tackled resorting to various iterative algorithms. Among them, one may mention the celebrated NONMEM software [8], which relies on maximum likelihood estimation, but also Bayesian algorithms that compute the posterior distribution of parameters exploiting the Markov chain Monte Carlo (MCMC) machinery [9], [10].

At the early stages of a study or when the mechanistic model of a physiological phenomenon is not available, it may be difficult to formulate a reliable parametric model. Hence the need for flexible nonparametric population approaches that reduce the structural assumptions to a minimum [11]. Along this direction, an example is provided by so-called semiparametric methods that model the response curves as regression splines [12], [13]. A potential difficulty underlying the use of these techniques is the optimization of the number and location of the knots of regression splines, which could suffer from the presence of local minima. More recently, in order to develop a fully nonparametric approach, within a Bayesian paradigm it has been proposed to model the individual curves as realizations of discrete- or continuous-time stochastic processes, e.g. random walks or integrated Wiener

* Corresponding author Gianluigi Pillonetto Ph. +390498277607

Email addresses: giapi@dei.unipd.it (Gianluigi Pillonetto), giuseppe.denicolao@unipv.it (Giuseppe De Nicolao), marloch@dei.unipd.it (Marco Chierici), cobelli@dei.unipd.it (Claudio Cobelli).

processes [14], [15]. In these works, each individual curve is seen as the sum of an average curve (common to all subjects) and an individual shift (varying from subject to subject). In particular, both the average curve and the individual shifts are assumed to be Gaussian processes whose statistics are specified by few hyper-parameters. For instance, if the curve is an integrated Wiener process, the hyperparameter is the corresponding intensity. Hyperparameter tuning can be carried out via likelihood maximization. For a given choice of the hyperparameters, the posterior expectations of the processes given the data provide point estimates of the average and individual curves. In particular, when the prior is formulated in terms of integrated Wiener processes, the estimated curves are cubic splines [15]. This Bayesian nonparametric approach has strong connections with kernel methods, Gaussian processes estimation, regularization networks [16],[17]. Recently, a Bayesian MCMC approach able to return the full posterior of hyperparameters and unknown functions has been also worked out, see [18].

In this paper, attention is focused on the nonparametric population analysis of standardized experiments which involve a large number of subjects. Herein, the term standardized is used to denote an experiment that is repeated in multiple subjects adopting a standard sampling schedule. A notable example, treated later in the paper, is the intravenous glucose tolerance test (IVGTT), where glucose plasma concentration is measured after intravenous administration of a glucose bolus. This test is widely employed in the diagnosis of metabolic disorders, see e.g. [19].

In the Bayesian nonparametric approach the computation of the posterior expectations calls for the solution of an algebraic linear system of order n_T , where n_T is the total number of observations. This is a potential drawback because the complexity scales with the cube of n_T . The burden may seem even worse in the case of standardized experiments involving a large number of subjects. As a matter of fact, in the present paper it is shown that the fixed sampling schedule can be exploited to design an algorithm whose complexity scales with the cube of the number of samples collected in each individual. This holds for evaluation of both the posterior expectations and confidence intervals.

The new algorithms pave the way to the implementation of a *client-server* architecture for managing the identification of population models for standardized experiments. The server computes and stores the sufficient statistics abstracted from a large historical data set. The client, whose aim is analyzing a single new experiment (not necessarily standardized), interrogates the server to get the information needed to compute the posterior expectation of the individual curve given all the historical data. The client can also send its data to the server in order to update the centralized sufficient statistics. As an example, the server could be managed by a reference research center, whereas the clients could be laboratories collecting and processing clinical data. According to this architecture, the local laboratories benefit from the information contained in the collective database in a computationally efficient way and without accessing individual data subject to privacy and confidentiality constraints. To the authors' knowledge, the client-server architecture is a novel

contribution of this paper. In fact, most population modeling approaches cannot be decentralized because of their intrinsic nonlinear-in-parameter structure.

The paper is organized as follows. In Section 2 the problem is given its mathematical formulation. In Section 3 the computational algorithms are derived. In Section 4 the proposed methodology is tested on a large data set of IVGTT experiments. Some conclusions end the paper.

2 Statement of the problem

In the sequel, $E[\cdot]$ is used to denote the expectation operator and vectors are column vectors, unless otherwise specified. Further, given two random vectors q and w , let $cov[q, w] = E[(q - E[q])(w - E[w])^T]$ and $Var[q] = E[(q - E[q])(q - E[q])^T]$.

We consider the problem of estimating realizations of continuous-time stochastic processes $x^j(t)$, $j = 1, 2, \dots, m+1$, from a finite number of noisy samples. The curves $x^j(t)$ represent the responses of $m+1$ subjects randomly drawn from a population. It is assumed that number and location of the sampling instants do not vary from subject to subject except for what concerns the last one. To be more specific, for $j = 1, \dots, m$ the curves are sampled at instants $\{t_k\}$, $k = 1, 2, \dots, n$, while the $(m+1)$ -th curve is sampled at instants $\{t_k^*\}$, $k = 1, 2, \dots, n^*$. The measurement model is

$$\begin{aligned} y_k^j &= x^j(t_k) + v_k^j, & k = 1, \dots, n, \quad j = 1, \dots, m \\ y_k^{m+1} &= x^{m+1}(t_k^*) + v_k^{m+1}, & k = 1, \dots, n^* \end{aligned}$$

where $\{v^j\} = [v_1^j \dots v_n^j]^T$, $j = 1, \dots, m$, and $\{v^{m+1}\} = [v_1^{m+1} \dots v_{n^*}^{m+1}]^T$ are Gaussian and independent random vectors such that for every k and j

$$E[v_k^j] = 0, \quad Var[v^j] = \Sigma_v^j$$

We assume that the individual curves can be decomposed as

$$x^j(t) = \bar{x}(t) + \tilde{x}^j(t), \quad j = 1, \dots, m+1$$

where $\bar{x}(t)$ and $\tilde{x}^j(t)$ are zero-mean normal stochastic processes that represent the average curve and the individual shift from the average, respectively. We also assume that processes $\{v_k^j\}_{j=1}^{m+1}$, $\bar{x}(t)$ and $\{\tilde{x}^j(t)\}_{j=1}^{m+1}$ are all mutually independent. For the sake of simplicity, it is assumed that $\{\tilde{x}^j(t)\}_{j=1}^{m+1}$ are identically distributed. Define now

$$\begin{aligned} y^j &= [y_1^j \quad y_2^j \quad \dots \quad y_n^j]^T, & j = 1, 2, \dots, m \\ y^{m+1} &= [y_1^{m+1} \quad y_2^{m+1} \quad \dots \quad y_{n^*}^{m+1}]^T \\ y &= [(y^1)^T \quad \dots \quad (y^m)^T]^T & y^+ = [y^T \quad (y^{m+1})^T]^T \end{aligned}$$

The paper is concerned with the solution of the following two estimation problems.

- Given y , for any t compute efficiently the continuous-time minimum variance estimate of the average curve $\bar{x}(t)$, i.e. $E[\bar{x}(t)|y]$, as well as the variance of the reconstruction error, i.e. $Var[\bar{x}(t)|y]$.
- Assuming that a new data set y^{m+1} is available, for any t compute efficiently $E[x^{m+1}(t)|y^+]$ and $Var[x^{m+1}(t)|y^+]$.

3 Computational algorithms

3.1 Computing $E[\bar{x}(t)|y]$ and $Var[\bar{x}(t)|y]$

The aim is to derive efficient algorithms to compute the estimates $E[\bar{x}(\tau)|y]$, $E[x^{m+1}(\tau)|y^+]$, where τ is a generic temporal instant, together with their confidence intervals. We start by introducing the following notation

$$\bar{x} = [\bar{x}(t_1) \dots \bar{x}(t_n)]^T \quad \bar{x}_\tau = [\bar{x}(\tau) \quad \bar{x}(t_1) \dots \bar{x}(t_n)]^T \quad (1)$$

$$\tilde{x}^j = [\tilde{x}^j(t_1) \quad \dots \quad \tilde{x}^j(t_n)]^T, \quad j = 1, 2, \dots, m$$

and

$$\begin{aligned} \bar{R} &= Var[\bar{x}] & \bar{R}_\tau &= cov[\bar{x}_\tau, \bar{x}] \\ \bar{R}_{\tau\tau} &= Var[\bar{x}_\tau] & \bar{r}_\tau &= cov[\bar{x}(\tau), \bar{x}] \\ \hat{R}_{\tau\tau} &= Var[\bar{x}_\tau|y] & \tilde{R} &= Var[\tilde{x}^j], \quad j = 1, 2, \dots, m \end{aligned} \quad (2)$$

Notice that in (2), due to the standardized sampling assumption, $\tilde{R} = Var[\tilde{x}^j]$ is independent of j . In the sequel, $bd\{A_1, \dots, A_p\}$ denotes the square block matrix whose diagonal elements are the square matrices A_1, \dots, A_p and the off-diagonal elements are zero. We also use I_d to denote the $d \times d$ identity matrix.

The following proposition exploits well known formulas regarding joint Gaussian vectors, see e.g. [20], [21].

Proposition 1 *Assuming that the autocovariances of $\bar{x}(t)$ and $\{\tilde{x}^j(t)\}_{j=1}^{m+1}$ are perfectly known, under the independence assumptions stated in Section 2, it holds that*

$$E[\bar{x}(\tau)|y] = cov[\bar{x}(\tau), y] Var[y]^{-1} y$$

$$\hat{R}_{\tau\tau} = \bar{R}_{\tau\tau} - cov[\bar{x}_\tau, y] Var[y]^{-1} cov[\bar{x}_\tau, y]^T \quad (3)$$

$$Var[y] = \begin{bmatrix} \bar{R} & \bar{R} & \dots & \bar{R} \\ \vdots & \vdots & \vdots & \vdots \\ \bar{R} & \bar{R} & \dots & \bar{R} \end{bmatrix} + bd\{\tilde{R} + \Sigma_v^1, \dots, \tilde{R} + \Sigma_v^m\} \quad (4)$$

In view of the above formulas, it would seem that the computation of $E[\bar{x}(\tau)|y]$ and $\hat{R}_{\tau\tau}$ calls for the inversion of an nm -th order matrix. The next proposition exploits the special structure of the problem to demonstrate that only n -th order inverses are required.

Proposition 2 *Let*

$$\tilde{R}_v^i = \tilde{R} + \Sigma_v^i, \quad F = \sum_{i=1}^m (\tilde{R}_v^i)^{-1}, \quad f = \sum_{i=1}^m (\tilde{R}_v^i)^{-1} y^i \quad (5)$$

Then,

$$\begin{aligned} \hat{R}_{\tau\tau} &= \bar{R}_{\tau\tau} - \bar{R}_\tau F \bar{R}_\tau^T + \bar{R}_\tau F [\bar{R}^{-1} + F]^{-1} F \bar{R}_\tau^T \\ E[\bar{x}(\tau)|y] &= \bar{r}_\tau c \end{aligned}$$

where $c \in \mathfrak{R}^n$ is given by $c = [I_n - F(\bar{R}^{-1} + F)^{-1}] f$.

Proof: In view of (4) and (5), we can rewrite $Var[y]$ as follows

$$Var[y] = \begin{bmatrix} \bar{R}^{\frac{1}{2}} & \dots & \bar{R}^{\frac{1}{2}} \end{bmatrix}^T \begin{bmatrix} \bar{R}^{\frac{1}{2}} & \dots & \bar{R}^{\frac{1}{2}} \end{bmatrix} + bd\{\tilde{R}_v^1, \dots, \tilde{R}_v^m\}$$

The matrix inversion lemma (e.g., see page 138 in [20])

$$(A + BCB^T)^{-1} = A^{-1} - A^{-1}B(C^{-1} + B^T A^{-1}B)^{-1}B^T A^{-1}$$

can now be used with the following assignments

$$A = bd\{\tilde{R}_v^1, \dots, \tilde{R}_v^m\} \quad B = \begin{bmatrix} \bar{R}^{\frac{1}{2}} & \dots & \bar{R}^{\frac{1}{2}} \end{bmatrix}^T \quad C = I_n$$

We obtain

$$\begin{aligned} Var[y]^{-1} &= bd\{(\tilde{R}_v^1)^{-1}, \dots, (\tilde{R}_v^m)^{-1}\} - \begin{bmatrix} (\tilde{R}_v^1)^{-1} \bar{R}^{\frac{1}{2}} \\ \vdots \\ (\tilde{R}_v^m)^{-1} \bar{R}^{\frac{1}{2}} \end{bmatrix} \\ &\quad \times [I_n + \sum_{i=1}^m \bar{R}^{\frac{1}{2}} (\tilde{R}_v^i)^{-1} \bar{R}^{\frac{1}{2}}]^{-1} \begin{bmatrix} \bar{R}^{\frac{1}{2}} (\tilde{R}_v^1)^{-1} & \dots & \bar{R}^{\frac{1}{2}} (\tilde{R}_v^m)^{-1} \end{bmatrix} \end{aligned}$$

which, using (5), can be rewritten as

$$\begin{aligned} Var[y]^{-1} &= bd\{(\tilde{R}_v^1)^{-1}, \dots, (\tilde{R}_v^m)^{-1}\} \\ &\quad - \begin{bmatrix} (\tilde{R}_v^1)^{-1} \\ \vdots \\ (\tilde{R}_v^m)^{-1} \end{bmatrix} [\bar{R}^{-1} + F]^{-1} \begin{bmatrix} (\tilde{R}_v^1)^{-1} & \dots & (\tilde{R}_v^m)^{-1} \end{bmatrix} \end{aligned} \quad (6)$$

Further, since $\bar{x}(t)$, $\{\tilde{x}^j(t)\}$ and $\{v^j\}$ are mutually independent, and the sampling schedule does not vary in the first m subjects, we have that

$$\begin{aligned} cov[\bar{x}_\tau, y] &= [cov[\bar{x}_\tau, y^1] \dots cov[\bar{x}_\tau, y^m]] \\ &= [cov[\bar{x}_\tau, \bar{x}] \dots cov[\bar{x}_\tau, \bar{x}]] = [\bar{R}_\tau \dots \bar{R}_\tau] \end{aligned}$$

Using (3), simple computations provide now the final expression for $\hat{R}_{\tau\tau}$.

As regards $E[\bar{x}(\tau)|y]$, we have that

$$E[\bar{x}(\tau)|y] = \left[\text{cov}[\bar{x}(\tau), y^1] \dots \text{cov}[\bar{x}(\tau), y^m] \right] \text{Var}[y]^{-1} y$$

Using (6),

$$\text{Var}[y]^{-1} y = (c^1 \dots c^m)^T \quad (7)$$

where

$$c^i = \left[(\tilde{R}_v^i)^{-1} y^i - (\tilde{R}_v^i)^{-1} (\bar{R}^{-1} + F)^{-1} \sum_{j=1}^m (\tilde{R}_v^j)^{-1} y^j \right]$$

Moreover, $\text{cov}[\bar{x}(\tau), y^i] = \text{cov}[\bar{x}(\tau), \bar{x}] c^i$, $i = 1, \dots, m$, so that

$$E[\bar{x}(\tau)|y] = \sum_{i=1}^m \text{cov}[\bar{x}(\tau), \bar{x}] c^i = \text{cov}[\bar{x}(\tau), \bar{x}] \sum_{i=1}^m c^i$$

where, using (5), one can easily see that $\sum_{i=1}^m c^i$ corresponds to vector c reported in the statement of the Proposition. This completes the proof. \square

3.2 Computing $E[x^{m+1}(t)|y^+]$ and $\text{Var}[x^{m+1}(t)|y^+]$

In the previous subsection, a method has been obtained to efficiently compute the estimate of the average curve of the population from data collected in m subjects on a standardized grid $\{t_k\}$. Assume now that such data are already available and that a new $(m+1)$ -th subject is sampled on an arbitrary sampling grid $\{t_k^*\}$. In this subsection, we show how to estimate the individual curve of the new subject taking into account all the available information on the previous m subjects. We start by reporting a notation which refers to the sampling grid $\{t_k^*\}$, and which represents an extension of that reported in (1) and (2). Let

$$\begin{aligned} \bar{x}^* &= [\bar{x}(t_1^*) \dots \bar{x}(t_n^*)]^T \\ \bar{x}_\tau^* &= [\bar{x}(\tau) \quad \bar{x}(t_1^*) \dots \bar{x}(t_n^*)]^T \\ \tilde{x}^{m+1} &= [\tilde{x}^{m+1}(t_1^*) \dots \tilde{x}^{m+1}(t_n^*)]^T \\ \tilde{R}_\tau^* &= \text{cov}[\bar{x}_\tau^*, \bar{x}], \quad \tilde{R}^* = \text{Var}[\tilde{x}^{m+1}] \\ \tilde{r}_\tau^* &= \text{cov}[\tilde{x}^{m+1}(\tau), \tilde{x}^{m+1}], \quad \tilde{r}_\tau = \text{Var}[\tilde{x}^{m+1}(\tau)] \\ \bar{R}_{\tau\tau}^* &= \text{Var}[\bar{x}_\tau^*], \quad \hat{R}_{\tau\tau}^* = \text{Var}[\bar{x}_\tau^*|y] \end{aligned}$$

The next proposition (whose proof is omitted) shows how to calculate efficiently $\hat{R}_{\tau\tau}^*$ and $E[\bar{x}_\tau^*|y]$.

Proposition 3 Consider quantities \tilde{R}_v^i, F and c as defined in Proposition 2. Then,

$$\begin{aligned} \hat{R}_{\tau\tau}^* &= \bar{R}_{\tau\tau}^* - \bar{R}_\tau^* F (\bar{R}_\tau^*)^T + \bar{R}_\tau^* F (\bar{R}^{-1} + F)^{-1} F (\bar{R}_\tau^*)^T \\ E[\bar{x}_\tau^*|y] &= \bar{R}_\tau^* c \end{aligned}$$

The last result allows one to compute the estimate of the average curve on an arbitrary sampling grid as well as confidence intervals thanks to the covariance matrix $\hat{R}_{\tau\tau}^*$. Notice that the formulas require only knowledge of the standardized sampling grid $\{t_k\}$, necessary in order to compute \bar{R}_τ^* and $\hat{R}_{\tau\tau}^*$, of matrix F and vector c . In some sense, F and c can be thought of as a sufficient statistic for average curve evaluation, given all the m subjects. In fact, once F and c are known, the individual data $\{y^j\}_{j=1}^m$ are no more necessary.

It is now shown how to efficiently compute $E[x^{m+1}(\tau)|y^+]$ as a function of $E[\bar{x}_\tau^*|y]$ and $\hat{R}_{\tau\tau}^*$. To obtain this goal, it is useful to partition $\hat{R}_{\tau\tau}^*$ and $E[\bar{x}_\tau^*|y]$ into submatrices as follows:

$$\begin{aligned} \hat{R}_{\tau\tau}^* &= \begin{bmatrix} \text{Var}[\bar{x}(\tau)|y] & \text{cov}[\bar{x}(\tau), \bar{x}^*|y] \\ \text{cov}[\bar{x}(\tau), \bar{x}^*|y]^T & \text{Var}[\bar{x}^*|y] \end{bmatrix} \doteq \begin{bmatrix} p & \bar{p} \\ \bar{p}^T & P \end{bmatrix} \quad (8) \\ E[\bar{x}_\tau^*|y] &= \begin{bmatrix} E[\bar{x}(\tau)|y] \\ E[\bar{x}^*|y] \end{bmatrix} \doteq \begin{bmatrix} \xi_\tau \\ \xi \end{bmatrix} \end{aligned}$$

A preliminary result is first needed.

Lemma 4 It holds that

$$\begin{aligned} E[y^{m+1}|y] &= E[\bar{x}^*|y] \\ \text{Var}[y^{m+1}|y] &= \text{Var}[\bar{x}^*|y] + \tilde{R}^* + \Sigma_v^{m+1} \\ E[x^{m+1}(\tau)|y] &= E[\bar{x}(\tau)|y] \\ \text{Var}[x^{m+1}(\tau)|y] &= \text{Var}[\bar{x}(\tau)|y] + \text{Var}[\tilde{x}^{m+1}(\tau)] \\ \text{cov}[x^{m+1}(\tau), y^{m+1}|y] &= \text{cov}[\bar{x}(\tau), \bar{x}^*|y] + \text{cov}[\tilde{x}^{m+1}(\tau), \tilde{x}^{m+1}] \end{aligned}$$

Proof To obtain the results, one can first replace y^{m+1} with $\bar{x}^* + \tilde{x}^{m+1} + v^{m+1}$ and $x^{m+1}(\tau)$ with $\bar{x}(\tau) + \tilde{x}^{m+1}(\tau)$. After that, one exploits the fact that \bar{x}^* , \tilde{x}^{m+1} and v^{m+1} are independent of each other and that \tilde{x}^{m+1} and v^{m+1} are independent of y . \square

We are now in a position to compute $E[x^{m+1}(\tau)|y^+]$ and the variance of the error affecting such estimate as a function of the submatrices of $\hat{R}_{\tau\tau}^*$ and $E[\bar{x}_\tau^*|y]$ given in (8). The following proposition exploits well known results on joint Gaussian vectors (e.g., see [20, Section 3.1]) and Lemma 4.

Proposition 5 It holds that

$$\begin{aligned} E[x^{m+1}(\tau)|y^+] &= \xi_\tau + \{\bar{p} + \tilde{r}_\tau^*\} (P + \tilde{R}^* + \Sigma_v^{m+1})^{-1} \\ &\quad \times (y^{m+1} - \xi) \\ \text{Var}[x^{m+1}(\tau)|y^+] &= p + \tilde{r}_\tau - \{\bar{p} + \tilde{r}_\tau^*\} \\ &\quad \times (P + \tilde{R}^* + \Sigma_v^{m+1})^{-1} \{\bar{p} + \tilde{r}_\tau^*\}^T \end{aligned}$$

The last result shows that the estimate of the $(m+1)$ -th individual curve depends only on the corresponding data y^{m+1} and matrix $\hat{R}_{\tau\tau}^*$ and vector $E[\bar{x}_\tau^*|y]$ in (8), which summarize all the information from the other m subjects. Recall that $\hat{R}_{\tau\tau}^*$

and $E[\bar{x}_t^*|y]$ depend only on F and c . Thus, one can imagine a distributed estimation scheme based on the following client-server architecture. The server processes m subjects sampled on a standardized grid $\{t_k\}$ and calculates matrix F and vector c . The client processes the data of the $(m+1)$ -th subject as follows: it asks the server for F and c and then uses them to obtain the individual curve. In this way, the client is able to exploit the whole data set made of m subjects without having access to the data.

Finally, notice also that, if we suppose that the $(m+1)$ -th subject has a standard sampling schedule, then its data can be sent to the server in order to update F and c . This will allow other clients to benefit from the information contained in y^{m+1} . It is worth stressing that such update operation can be performed very efficiently. In fact, denote with F^u and c^u the updated versions of F and c , respectively. Then, exploiting the results reported in Proposition 2, the following update formulas easily follow

$$F^u = F + (\tilde{R}_v^{m+1})^{-1} \quad c^u = \left[I_n - F^u (\bar{R}^{-1} + F^u)^{-1} \right] f^u$$

$$f^u = f + (\tilde{R}_v^{m+1})^{-1} y^{m+1}$$

3.3 Computing unknown hyper-parameters

When dealing with real world problems, the autocovariances of the signals of interest may contain unknown hyper-parameters which have to be estimated from data. To tackle this problem, we resort to the so-called Empirical Bayes method where a Maximum Likelihood (ML) estimate of hyper-parameters is first achieved. Next, parameters are set to their ML estimates and the Bayes estimate is computed using the formulas described in the previous sub-sections. Within the client-server architecture outlined at the end of the previous subsection, it is the server that is in charge of estimating the hyperparameters and transmitting them to the clients. In fact, the client needs the hyperparameters in order to be able to evaluate the autocovariances which enter the computational formulas. Now, assume that the unknown hyper-parameters are contained in the parameter vector θ that parametrizes $Var[y]$. To compute their estimates, the following optimization problem has to be solved

$$\hat{\theta} = \arg \min_{\theta} J(y; \theta)$$

where $J(y; \theta) = \log[\det(Var[y])] + y^T Var[y]^{-1} y$ is equal to the opposite of the log-likelihood apart from a constant. In the following two propositions, it is shown that evaluation of the log-likelihood for a given θ requires only $O(m \times n^3)$ operations. The first result can be immediately deduced by exploiting (7).

Proposition 6 *It holds that*

$$y^T Var[y]^{-1} y = y^T c$$

In the sequel, we use $chol[K]$ to denote the Cholesky factorization of the matrix K [20]. Let also $A_{kk} = \bar{R} + \tilde{R} + \Sigma_v^k$ indicate the k -th $n \times n$ block on the diagonal of $Var[y]$. The second result (whose proof is omitted) is reported below.

Proposition 7 *The determinant of the $nm \times nm$ matrix $Var[y]$ can be computed as*

$$\det(Var[y]) = \left(\prod_{i=1}^m \det(C_{ii}) \right)^2$$

where C_{ii} are $n \times n$ matrices such that

$$C_{ii} = chol[A_{ii} - D_i], \quad C_{(i+1)i} = (\bar{R} - D_i) (C_{ii}^T)^{-1}$$

$$C_{ki} = C_{(i+1)i} \text{ for } k > i + 1, \quad D_{i+1} = D_i + C_{(i+1)i} C_{(i+1)i}^T$$

with the initial positions $D_1 = 0$ and $C_{21} = \bar{R} (C_{11}^T)^{-1}$.

4 Example

In this Section, the proposed computational scheme is tested on a database that consists of 224 healthy subjects, on which an intravenous glucose tolerance test (IVGTT) was performed. The first 204 subjects represent our training set. It consists of standardized experiments where the same sampling schedule was adopted for all subjects and plasma glucose samples were collected non-uniformly between 0 and 240 min [22]. To be specific, $n = 20$ and the set Ω_1 , expressed in minutes, corresponds to the sampling instants $\{t_k\}$ given by $\{2, 4, 6, 8, 10, 15, 20, 22, 25, 26, 28, 31, 35, 45, 60, 75, 90, 120, 180, 240\}$. Data regarding the remaining 20 subjects are selected from IVGTT studies described in [5,23,24], in which a glucose dose, identical to that administered in the first 204 subjects, was injected at time zero. In particular, in the first 13 subjects $n = 30$ and the sampling grid, denoted by Ω_1 and expressed in minutes, is $\{2, 3, 4, 5, 6, 8, 10, 12, 15, 20, 22, 24, 26, 28, 30, 35, 40, 45, 50, 55, 60, 70, 80, 100, 120, 140, 160, 180, 210, 240\}$. In the last 7 subjects $n = 29$ and the sampling grid, denoted by Ω_2 , is $\{2, 3, 4, 5, 8, 10, 12, 14, 16, 18, 20, 24, 28, 32, 40, 45, 50, 60, 70, 80, 90, 100, 110, 120, 140, 160, 180, 210, 240\}$. In all the 224 subjects, measurements are known to be corrupted by a white normal noise with a 2% coefficient of variation. Glucose data were pre-processed by first subtracting the basal value from each profile. To take into account the fact that processed glucose data tend to zero, a time transformation of t was performed so that the new time \tilde{t} ranges from 0 to 1, according to the formula $\tilde{t} = 1/(1+t/\gamma)$, as in [15]. We set parameter γ to 30, a value that maximizes the minimum distance between each pair of transformed sampling instants given by the union of the sampling grids associated with the 224 subjects. Both the typical curve and the individual shifts were modelled as integrated Wiener processes.

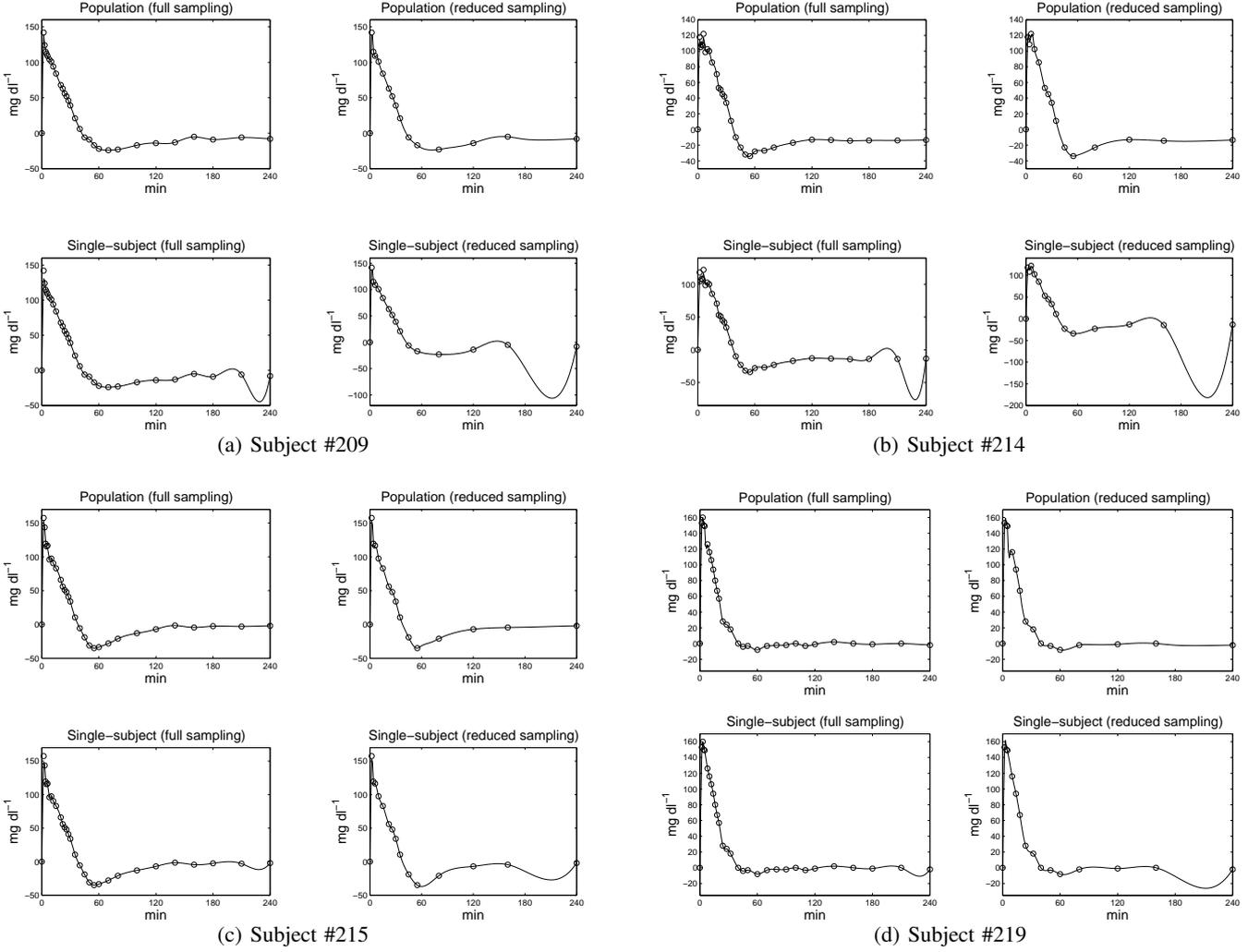


Fig. 2. Estimation of individual IVGTT responses. Comparison between Population approach and Single-subject approach in 4 representative subjects. The Population approach exploits information from other 204 IVGTT experiments.

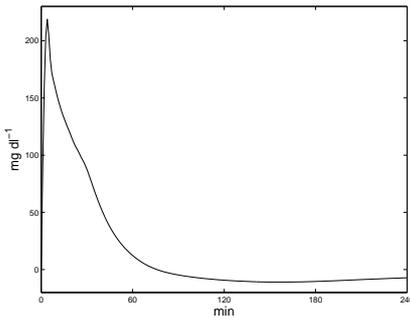


Fig. 1. Estimated average curve obtained by Population approach applied to 204 IVGTT responses

Thus, it holds that (see also [15])

$$\text{cov}(\bar{x}(s), \bar{x}(\tau)) = \bar{\lambda}^2 \begin{cases} \frac{s^2}{2} (\tau - \frac{s}{3}) & s \leq \tau \\ \frac{\tau^2}{2} (s - \frac{\tau}{3}) & s > \tau \end{cases}$$

$$\text{cov}(\tilde{x}^j(s), \tilde{x}^j(\tau)) = \tilde{\lambda}^2 \begin{cases} \frac{s^2}{2} (\tau - \frac{s}{3}) & s \leq \tau \\ \frac{\tau^2}{2} (s - \frac{\tau}{3}) & s > \tau \end{cases} \quad \forall j$$

As previously observed, hyperparameters $\bar{\lambda}$ and $\tilde{\lambda}$ have to be estimated from data.

Data relative to the first $m = 204$ subjects were exploited to estimate hyper-parameters via maximum likelihood. One can think of the computations related to such data base as e.g. performed by a reference research center which has access to a large data set. Obtained values for $\bar{\lambda}$ and $\tilde{\lambda}$ turned out to be around 13700 and 4440, respectively. After that, each of the remaining 20 “ $(m + 1)$ -th” curves, was estimated by applying the method discussed in Section 3, hereby called “population approach”. For comparison, the “ $(m + 1)$ -th” profiles were also estimated without exploiting information regarding the first 204 subjects. In particular, curves were modeled as integrated Wiener processes with regularization parameter estimated via maximum likelihood and adopting the same time transformation \tilde{t} . Note that, under the integrated Wiener prior, the estimated curves are cubic smoothing splines [25]. This approach will be hereby called “single-

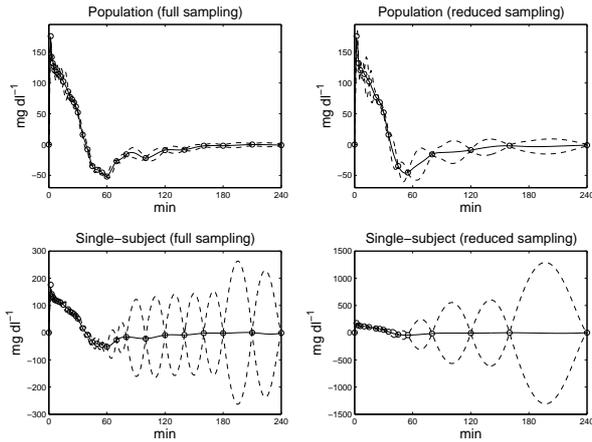


Fig. 3. Results obtained in subject #211. Dashed lines denote $\pm 1SD$ confidence bands

subject approach”. As concerns measurements related to the “ $(m+1)$ -th” subject, we employed both a full and reduced sampling schedule. In the latter, when dealing with the sampling grid Ω_1 , n^* is fixed to 15, and the set $\{t_k^*\}$ corresponds to $\{2, 4, 6, 10, 15, 22, 26, 30, 35, 45, 55, 80, 120, 160, 240\}$. When dealing with the sampling grid Ω_2 , n^* is still 15 but the set $\{t_k^*\}$ now corresponds to $\{2, 4, 5, 10, 14, 18, 24, 32, 40, 50, 60, 80, 120, 160, 240\}$. One can think of computations related to the “ $(m+1)$ -th” curves as performed by clients such as laboratories collecting clinical data.

Fig. 1 plots the estimated average curve. The results regarding 4 representative subjects (#209, #214, #215 and #219) are instead visible in Fig.2 where t is used as time axis. In each of the sub-figures, fully sampled data vs. estimated curves are plotted for the population approach (top-left) and the single-subject one (bottom-left). The remaining two panels in each sub-figure show reduced data vs estimated curves for the population approach (top-right) and single-subject one (bottom-right). Rather interestingly, the curves reconstructed via the single-subject approach suffer from oscillations in the final part of the experiment ($t > 120$ min), where data are sampled less frequently. On the other hand, the population approach is less influenced by the sampling schedule. In fact, good results are obtained with both full and reduced sampling. In Fig. 3 we show results regarding subject #211 reporting estimated curves and $\pm 1SD$ confidence bands (dashed lines). It is apparent that confidence intervals obtained by the population approach are narrower than those obtained by the single-subject approach. To obtain a global comparison of the two approaches, given the full and reduced sampling grids I_f and I_r , respectively, define $I = I_f \setminus I_r$. Then, for every “ $(m+1)$ -th” subject we compute the quantity

$$RMSE = \sqrt{\frac{\sum_{t \in I} (y^{m+1}(t) - \hat{x}(t))^2}{n - n^*}}$$

where $\hat{x}(t) = E[x^{m+1}(t)|y^+]$. Notice that the smaller RMSE, the larger the predictive capability at times falling in the full

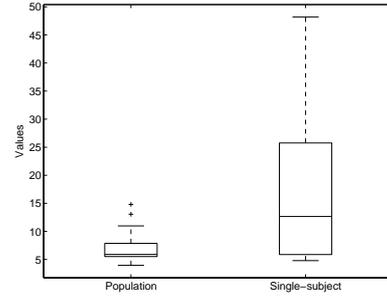


Fig. 4. Estimation of individual IVGTT responses. Comparison of RMSE obtained by Population approach and Single-subject approach in 20 subjects

sampling grid but not used in the estimation process. Fig. 4 plots boxplots of the 20 computed values of RMSE using the population and single subject approach. It is apparent that the population approach outperforms the single-subject one in terms of predictive capability on new glucose data.

5 Conclusions

In this paper, efficient formulas have been developed for the nonparametric estimation of population models from large data sets collected in experiments adopting a standardized sampling schedule. In particular, the structure of the problem has been exploited to design algorithms whose complexity scales with the cube of the number of data collected in a single subject, while previous algorithms [15,18] scaled with the cube of the total number of observations. The overall identification scheme presents a “client-server” architecture. The server takes care of managing historical information on past experiments. The client deals with a single new experiment and interrogates the server to obtain the information needed to reconstruct the individual curve. In this way, clients exploit the global data set without having access to the historical data and with negligible computational effort. A data base of 224 IVGTT experiments has been employed to successfully validate the approach. Obtained results show that the individual estimates provided by the population approach are much more satisfactory than those achievable from the single experiment. In addition, when the sampling schedule of a subject is reduced, the individual estimate obtained by the population approach remains satisfactory whereas the quality of the single-subject estimate shows a further degradation.

Acknowledgments

We thank Dr. R.A. Rizza and Dr. R. Basu (Mayo Clinic, Rochester, MN USA) for having made available to us the IVGTT data of [22]. Data described in [5,23,24] were downloaded from the website of the Resource Facility for Population Kinetics (<http://www.rfpk.washington.edu>). This research has been partially supported by FIRB Project “Learning theory and application”, by NIH/NIBIB grant P41-EB01975 and by the PRIN Project “New Methods

and Algorithms for Identification and Adaptive Control of Technological Systems”.

References

- [1] S. L. Beal and L. B. Sheiner. Estimating population kinetics. *Crit. Rev. Biomed. Eng.*, 8(3):195–222, 1982.
- [2] L. B. Sheiner. The population approach to pharmacokinetic data analysis: rationale and standard data analysis methods. *Drug Metabolism Reviews*, 15:153–171, 1994.
- [3] M. Davidian and D. M. Giltinan. *Nonlinear Models for Repeated Measurement Data*. Chapman and Hall, New York, 1995.
- [4] A. Bertoldo, G. Sparacino, and C. Cobelli. “Population” approach improves parameter estimation of kinetic models from dynamic PET data. *IEEE Trans. on Medical Imaging*, 23(3):297–306, 2004.
- [5] P. Vicini and C. Cobelli. The iterative two-stage population approach to IVGTT minimal modeling: improved precision with reduced sampling. *Am. J. Physiol. Endocrinol. Metab.*, 280(1):179–186, 2001.
- [6] F. Ferrazzi, P. Magni, and R. Bellazzi. Bayesian clustering of gene expression time series. In *Proc. of 3rd Int. Workshop on Bioinformatics for the Management, Analysis and Interpretation of Microarray Data (NETTAB 2003)*, pages 53–55, 2003.
- [7] J.A. Jacquez. *Compartmental analysis in biology and medicine*. Ann Arbor: University of Michigan Press, 1985.
- [8] S. Beal and L. Sheiner. *NONMEM User’s Guide*. NONMEM Project Group, University of California, San Francisco, 1992.
- [9] W.R. Gilks, S. Richardson, and D.J. Spiegelhalter. *Markov chain Monte Carlo in Practice*. London: Chapman and Hall, 1996.
- [10] D. J. Lunn, N. Best, A. Thomas, J. C. Wakefield, and D. Spiegelhalter. Bayesian analysis of population PK/PD models: general concepts and software. *J. Pharmacokin. Pharmacodyn.*, 29(3):271–307, 2002.
- [11] I.A. Ibragimov and R.Z. Khasminskii. *Statistical Estimation: Asymptotic Theory*. Springer, 1981.
- [12] K. E. Fattinger and D. Verotta. A nonparametric subject-specific population method for deconvolution: I. Description, internal validation, and real data examples. *J. Pharmacokin. Biopharm.*, 23:581–610, 1995.
- [13] K. Park, D. Verotta, T. F. Blaschke, and L. B. Sheiner. A semiparametric method for describing noisy population pharmacokinetic data. *J. Pharmacokin. Biopharm.*, 25:615–642, 1997.
- [14] P. Magni, R. Bellazzi, G. De Nicolao, I. Poggesi, and M. Rocchetti. Nonparametric AUC estimation in population studies with incomplete sampling: a Bayesian approach. *J. Pharmacokin. Pharmacodyn.*, 29(5/6):445–471, 2002.
- [15] M. Neve, G. De Nicolao, and L. Marchesi. Nonparametric identification of population models via Gaussian processes. *Automatica*, 43:1134–1144, 2007.
- [16] C.E. Rasmussen and C.K.I. Williams. *Gaussian Processes for Machine Learning*. The MIT Press, 2006.
- [17] T. Evgeniou, C.A. Micchelli, and M. Pontil. Learning multiple tasks with kernel methods. *J. Machine Learning Research*, 6:615–637, 2005.
- [18] M. Neve, G. De Nicolao, and L. Marchesi. Nonparametric identification of population models: An MCMC approach. *IEEE Trans. on Biomedical Engineering*, 55:41–50, 2008.
- [19] R.N. Bergman, C.R. Bowden, and C. Cobelli. *Carbohydrate Metabolism*, chapter The minimal model approach to quantification of factors controlling glucose disposal in man. Wiley, New York, 1981.
- [20] B. D. O. Anderson and J. B. Moore. *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, N.J., USA, 1979.
- [21] A. N. Shiryaev. *Probability*. Springer, New York, NY, USA, 1996.
- [22] R. Basu, C. Dalla Man, M. Campioni, A. Basu, G. Klee, G. Jenkins, G. Toffolo, C. Cobelli, and R.A. Rizza. Effect of age and sex on postprandial glucose metabolism: difference in glucose turnover, insulin secretion, insulin action, and hepatic insulin extraction. *Diabetes*, 55:2001–2014, 2006.
- [23] P. Vicini, A. Caumo, and C. Cobelli. The hot IVGTT two-compartment minimal model: indexes of glucose effectiveness and insulin sensitivity. *Am. J. Physiol.*, 273:1024–1032, 1997.
- [24] P. Vicini, J.J. Zachwieja, K.E. Yarasheski, D.M. Bier, A. Caumo, and C. Cobelli. Glucose production during an IVGTT by deconvolution: validation with the tracer-to-tracee clamp technique. *Am. J. Physiol.*, 276:285–294, 1999.
- [25] G. Wahba. *Spline Models for Observational Data*. SIAM, Philadelphia, 1990.