

Nonparametric identification of population pharmacokinetic models: An MCMC approach

Marta Neve, Giuseppe De Nicolao, Laura Marchesi
Dipartimento di Informatica e Sistemistica
Università di Pavia, via Ferrata 1, 27100 Pavia, Italy
e-mail: marta.neve@unipv.it, giuseppe.denicolao@unipv.it

Abstract—The paper deals with the nonparametric identification of population models, that is models that explain jointly the behaviour of different subjects drawn from a population, e.g. responses of different patients to a drug. The average response of the population and the individual responses are modelled as continuous-time Gaussian processes with unknown hyperparameters. The posterior expectation and variance of both the average and individual curves are computed by means of a Markov Chain Monte Carlo scheme. The model and the estimation procedure are tested on xenobiotics pharmacokinetic data.

I. INTRODUCTION

In science and technology, it is rather common to analyze data coming from multiple experiments performed on different subjects belonging to some given population. A typical example is found in biomedical data analysis, where different subjects are sampled and analyzed in order to assess not only an average model but also to characterize the inter-individual variability within the population. For instance, population analysis is routinely used in the drug development process to estimate pharmacokinetic and pharmacodynamic models.

In the most favourable case, enough data are available for each subject so as to be able to reliably identify an individual model. Then, the population can be characterized by the distribution of the individual models. In some cases, however, the number of observations that can be collected in each subject is limited by technical, ethical and cost reasons. In pharmacological experiments this happens for toxicokinetic studies or the studies carried on populations that include “critical” subjects such as neonatal or pediatric or intensive care unit patients. When it is not possible to estimate each individual model separately, so-called “population methods” must be used. Such methods analyze all the data jointly, yielding the average and, possibly, also the individual models.

A first class of population methods relies on parametric models, usually formulated as grey-box continuous-time dynamical systems [1], [2], [3], [4], [5] (note that the term “nonparametric” in the last two papers refers to the estimation of the probability distributions of the parameters of a grey-box model). In many cases the physical knowledge may not suffice to specify a structural model. For such a reason semi-parametric and nonparametric population models have been proposed. The semi-parametric model [6]

relies on regression splines so that the nontrivial problem of deciding the number and the locations of the spline knots arises. As an alternative, nonparametric models that describe the individual curves as realizations of a discrete-time [7] or continuous-time [8] stochastic process have been proposed. Recently, a nonparametric model of this type has also been employed to analyze gene expression time series measured using DNA micro-arrays [9]. In [8], model identification is performed according to an Empirical Bayes approach that, though relatively simple, provides only point estimates of the so-called hyperparameters and, as such, underestimates the confidence intervals. The novelty of the present paper is that it develops a fully Bayesian identification scheme, which resorts to a Markov Chain Monte Carlo (MCMC) procedure for estimating the posterior expectations and variances. The method is tested on pharmacokinetic data related to xenobiotics administration in human subjects.

II. THE POPULATION MODEL

Nonparametric identification of population models calls for the estimation of an entire family of scalar real-valued continuous-time functions starting from a set of noisy samples taken at discrete time instants. Each function represents the behaviour of a subject belonging to the population of interest, e.g. it could be an impulse response such as the profile of drug concentration in plasma after administration of a unit bolus. In the following, $z^j(t)$, $j = 1, \dots, N$, $t \geq 0$, will denote the response curve associated with the j -th subject, while the available measurements are

$$y_k^j = z^j(t_k^j) + v_k^j, \quad k = 1, \dots, n_j, \quad (1)$$

where $t_k^j > 0$ denotes the k -th sampling instant (“knot”) for the j -th curve, and the measurements errors v_k^j are mutually independent and normally distributed with $E[v_k^j] = 0$, $Var[v_k^j] = (\sigma_k^j)^2$. The overall number of data will be indicated by $n = n_1 + n_2 + \dots + n_N$. It is worth noting that the number and location of the sampling instants t_k^j may vary from subject to subject. It is assumed that each individual curve can be decomposed as

$$z^j(t) = \bar{z}(t) + \tilde{z}^j(t)$$

where $\bar{z}(t)$ is the “typical curve” (also called average curve) of the population and $\tilde{z}^j(t)$ is the “individual shift” with

respect to the average behaviour. For ease of notation, it is convenient to define the following vectors

$$\begin{aligned} \mathbf{y} &:= [y_1^1 \dots y_{n_1}^1 \ y_1^2 \dots y_{n_2}^2 \dots y_1^N \dots y_{n_N}^N]^T \\ \bar{\mathbf{z}} &:= [\bar{z}(t_1^1) \dots \bar{z}(t_{n_1}^1) \ \dots \bar{z}(t_1^N) \dots \bar{z}(t_{n_N}^N)]^T \\ \tilde{\mathbf{z}} &:= [\tilde{z}^1(t_1^1) \dots \tilde{z}^1(t_{n_1}^1) \ \dots \tilde{z}^N(t_1^N) \dots \tilde{z}^N(t_{n_N}^N)]^T \\ \mathbf{v} &:= [v_1^1 \dots v_{n_1}^1 \ v_1^2 \dots v_{n_2}^2 \dots v_1^N \dots v_{n_N}^N]^T \end{aligned}$$

so that (1) can be rewritten as

$$\mathbf{y} = \bar{\mathbf{z}} + \tilde{\mathbf{z}} + \mathbf{v} \quad (2)$$

where $\mathbf{v} \sim N(0, \Sigma_v)$, $\Sigma_v := \text{diag}\{(\sigma_1^1)^2 \dots (\sigma_{n_N}^N)^2\}$.

A. Continuous-time model

In the present paper a stochastic approach is adopted: the typical curve and the individual shifts are assumed to be realizations of continuous-time Gaussian processes.

Assumption 1: The Gaussian stochastic processes $\bar{z}(t)$ and $\tilde{z}^j(t)$, $j = 1, \dots, N$, are independent of each other and of the noise samples \mathbf{v} . \square

Regarding the statistics of the signals, it is natural to assume that the individual shifts with respect to the average behaviour have zero mean, i.e. $E[\tilde{z}^j(t)] = 0, \forall t, \forall j$. Moreover, by properly normalizing the data, it can be assumed without loss of generality that $E[\bar{z}(t)] = 0$ as well. In the following $\bar{R}(t, \tau) := \text{Cov}[\bar{z}(t), \bar{z}(\tau)]$ and $\tilde{R}(t, \tau) := \text{Cov}[\tilde{z}^j(t), \tilde{z}^j(\tau)]$, $\forall j$, will denote the autocovariance functions of the typical curve and the individual shifts, respectively.

In a stochastic approach the signal model reflects the available a-priori knowledge. For instance, if it is only known that a signal is “smooth”, the following integrated Wiener process can be used.

Assumption 2:

$$\begin{aligned} \dot{\bar{x}}(t) &= \bar{\mathbf{A}}\bar{x}(t) + \bar{\mathbf{B}}\bar{w}(t) \\ \bar{z}(t) &= \bar{\mathbf{C}}\bar{x}(t) \end{aligned}$$

$$\bar{\mathbf{A}} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad \bar{\mathbf{B}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \bar{\mathbf{C}} = [1 \ 0]$$

where $\bar{x}(0) \sim N(0, \bar{\lambda}^2 \bar{\mathbf{X}}_0)$, and $\bar{w}(t)$ is a scalar continuous-time white Gaussian noise, independent of $\bar{x}(0)$ and \mathbf{v} , with $E[\bar{w}(t)\bar{w}(\tau)] = \bar{\lambda}^2 \delta(t - \tau)$. \square

The matrix $\bar{\mathbf{X}}_0$ is assumed to be known, whereas the parameter $\bar{\lambda}^2$ acts as a tuning knob: smaller values correspond to smoother signals. Since it is often difficult to specify a proper value for $\bar{\lambda}^2$ on the basis of prior knowledge alone, $\bar{\lambda}^2$ is regarded as a “hyperparameter” that has to be estimated from the data. As is well known, the autocovariance $\bar{R}(t, \tau)$ has the following expression:

$$\bar{R}(t, \tau) = \bar{\lambda}^2 \bar{\gamma}(t, \tau)$$

$$\bar{\gamma}(t, \tau) = \begin{cases} \bar{\mathbf{C}}\bar{\mathbf{X}}(t)e^{\bar{\mathbf{A}}^T(\tau-t)}\bar{\mathbf{C}}^T & t \leq \tau \\ \bar{\mathbf{C}}e^{\bar{\mathbf{A}}(t-\tau)}\bar{\mathbf{X}}(\tau)\bar{\mathbf{C}}^T & t > \tau \end{cases}$$

where $\bar{\mathbf{X}}(t) = \text{Var}[\bar{x}(t)]/\bar{\lambda}^2$ is the solution of the differential Lyapunov equation

$$\begin{aligned} \dot{\bar{\mathbf{X}}}(t) &= \bar{\mathbf{A}}\bar{\mathbf{X}}(t) + \bar{\mathbf{X}}(t)\bar{\mathbf{A}}^T + \bar{\mathbf{B}}\bar{\mathbf{B}}^T \\ \bar{\mathbf{X}}(0) &= \bar{\mathbf{X}}_0. \end{aligned}$$

It turns out that, for a given value $\tau = \bar{\tau}$, the function $\bar{\gamma}(t, \bar{\tau})$ is a piecewise cubic polynomial and is continuous with all its derivatives everywhere but in $t = \bar{\tau}$, where it is continuous up to its second derivative. It would be tempting to use the integrated Wiener model to describe also the individual shifts $\tilde{z}^j(t)$. However, as explained elsewhere [8], it is more convenient to model the individual shifts as stochastic processes whose variance does not grow unbounded as t goes to infinity.

Assumption 3: For $j = 1, \dots, N$,

$$\begin{aligned} \dot{\tilde{x}}(t) &= \tilde{\mathbf{A}}\tilde{x}(t) + \tilde{\mathbf{B}}\tilde{w}^j(t) \\ \tilde{z}^j(t) &= \tilde{\mathbf{C}}\tilde{x}(t) \end{aligned}$$

$$\tilde{\mathbf{A}} = \begin{bmatrix} a_1 & 1 \\ 0 & a_2 \end{bmatrix}, \quad \tilde{\mathbf{B}} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}, \quad \tilde{\mathbf{C}} = [1 \ 0]$$

where $a_1 < 0, a_2 < 0$, $\tilde{x}(0) \sim N(0, \tilde{\lambda}^2 \tilde{\mathbf{X}}_0)$, and $\tilde{w}^j(t)$ is a scalar continuous-time white Gaussian noise (independent of $\tilde{x}(0)$, \mathbf{v} , $\bar{w}(t)$ and $\tilde{w}^i(t)$, $i \neq j$) with $E[\tilde{w}^j(t)\tilde{w}^j(\tau)] = \tilde{\lambda}^2 \delta(t - \tau)$. \square

The autocovariance function of $\tilde{z}^j(t)$ is obtained as $\tilde{R}(t, \tau) = \tilde{\lambda}^2 \tilde{\gamma}(t, \tau)$, where $\tilde{\gamma}(t, \tau)$ is computed in an analogous way as $\bar{\gamma}(t, \tau)$. Assuming that $\tilde{\mathbf{X}}_0$ is known, the statistics of $\tilde{z}(t)$ depends on the three parameters $a_1, a_2, \tilde{\lambda}^2$. For $\tilde{\lambda}^2$ the same considerations as for $\bar{\lambda}^2$ hold. The two poles a_1 and a_2 provide a few more degrees of freedom for shaping the spectrum of $\tilde{z}^j(t)$. When the number of data per subject is scarce, it may be difficult to reliably estimate a_1 and a_2 , in which case their values are fixed according to the available prior knowledge.

B. Sampled model

In this section, a sampled reformulation of the continuous-time model is introduced that will prove useful in the following. More precisely, (2) is rewritten as

$$\mathbf{y} = \Phi \bar{\mathbf{w}} + \mathbf{D}_a \tilde{\mathbf{w}} + \mathbf{v} \quad (3)$$

where Φ and \mathbf{D}_a are such that $\bar{\lambda}^2 \Phi \Phi^T = \text{Var}[\bar{\mathbf{z}}]$, $\tilde{\lambda}^2 \mathbf{D}_a \mathbf{D}_a^T = \text{Var}[\tilde{\mathbf{z}}]$, $\bar{\mathbf{w}} \sim N(0, \bar{\lambda}^2 \mathbf{I}_n)$ and $\tilde{\mathbf{w}} \sim N(0, \tilde{\lambda}^2 \mathbf{I}_n)$. It is easy to see that:

$$\text{Var}[\bar{\mathbf{z}}] = \bar{\lambda}^2 \begin{bmatrix} \bar{\gamma}(t_1^1, t_1^1) & \dots & \bar{\gamma}(t_1^1, t_{n_N}^N) \\ \dots & \dots & \dots \\ \bar{\gamma}(t_{n_N}^N, t_1^1) & \dots & \bar{\gamma}(t_{n_N}^N, t_{n_N}^N) \end{bmatrix}$$

$$\text{Var}[\bar{\mathbf{z}}] = \text{blockdiag}\{\tilde{\mathbf{R}}^1, \dots, \tilde{\mathbf{R}}^N\}$$

$$\tilde{\mathbf{R}}^j := \tilde{\lambda}^2 \begin{bmatrix} \tilde{\gamma}(t_1^j, t_1^j) & \dots & \tilde{\gamma}(t_1^j, t_{n_j}^j) \\ \dots & \dots & \dots \\ \tilde{\gamma}(t_{n_j}^j, t_1^j) & \dots & \tilde{\gamma}(t_{n_j}^j, t_{n_j}^j) \end{bmatrix}$$

In the following, the values of the poles, a_1 and a_2 , (entering the definition of the matrix \mathbf{D}_a) as well as the noise variance, Σ_v , are supposed to be known, so that only two hyperparameters have to be estimated from the data: $\tilde{\lambda}^2$ and $\tilde{\lambda}^2$.

III. MCMC ESTIMATION OF THE SAMPLED MODEL

Consider a generic identification problem in which \mathbf{y} denotes the observed data and θ the unknown parameters of the model. Bayesian inference requires the computation of the posterior distribution $p(\theta|\mathbf{y})$. Given $p(\theta|\mathbf{y})$, a point estimate of θ is

$$E[\theta|\mathbf{y}] = \int \theta p(\theta|\mathbf{y}) d\theta.$$

In our problem, this expression does not admit an analytical solution and its numerical evaluation is tackled by means of an MCMC approach.

A. MCMC estimation

Monte Carlo integration evaluates $E[\theta|\mathbf{y}]$ by drawing samples $\{\theta_i, i = 1, \dots, h\}$ from the posterior distribution (“full conditional distribution”) so that

$$E[\theta|\mathbf{y}] \approx \frac{1}{h} \sum_{i=1}^h \theta_i.$$

It is not necessary that the samples are drawn independently from $p(\theta|\mathbf{y})$, but it suffices that they explore the whole support of $p(\theta|\mathbf{y})$ in the correct proportions. One way of doing this is through a Markov chain having $p(\theta|\mathbf{y})$ as its stationary distribution. This result can be obtained resorting to a well known family of algorithms named Metropolis-Hastings algorithms [10]. In particular, letting $\theta = [\tilde{\lambda}^2 \tilde{\lambda}^2]^T$, the Gibbs sampler will be applied to the estimation of the population model previously described.

B. Hyperparameters priors

The very first step consists of assigning a prior distribution to all the random variables involved in the model. A slightly informative prior is to be preferred for parameters that are affected by a large uncertainty. For the hyperparameters $\tilde{\lambda}^2$ and $\tilde{\lambda}^2$, a computationally advantageous choice, which simplifies the subsequent calculation of the so-called full conditional distributions, is to model the prior of $1/\tilde{\lambda}^2$ and $1/\tilde{\lambda}^2$ as a Gamma distribution with large (possibly infinite) variance:

$$p\left(\frac{1}{\tilde{\lambda}^2}\right) = \Gamma(g_1, g_2) \propto \left(\frac{1}{\tilde{\lambda}^2}\right)^{g_1-1} e^{-\frac{g_2}{\tilde{\lambda}^2}}$$

$$p\left(\frac{1}{\tilde{\lambda}^2}\right) = \Gamma(g_3, g_4) \propto \left(\frac{1}{\tilde{\lambda}^2}\right)^{g_3-1} e^{-\frac{g_4}{\tilde{\lambda}^2}}$$

where g_i are the parameters that characterize the Gamma distribution.

C. Full conditional distributions

The full conditional is the probability distribution of a variable conditioned on all the other variables in the model. It turns out that the full conditional for $1/\tilde{\lambda}^2$ is $\Gamma(\bar{g}_1, \bar{g}_2)$ with arguments $\bar{g}_1 = g_1 + n/2$ and $\bar{g}_2 = g_2 + \bar{\mathbf{w}}^T \bar{\mathbf{w}}/2$. As for the full conditional of $1/\tilde{\lambda}^2$, it is $\Gamma(\bar{g}_3, \bar{g}_4)$, $\bar{g}_3 = g_3 + n/2$, $\bar{g}_4 = g_4 + \tilde{\mathbf{w}}^T \tilde{\mathbf{w}}/2$. The full conditional distributions of $\bar{\mathbf{w}}$ and $\tilde{\mathbf{w}}$ are proved to be multivariate Gaussian functions characterized by the following statistics:

$$\text{Var}[\bar{\mathbf{w}}|\cdot] = \left(\tilde{\lambda}^{-2} \mathbf{I}_n + \Phi^T (\text{Var}[\bar{\mathbf{z}}] + \Sigma_v)^{-1} \Phi\right)^{-1}$$

$$E[\bar{\mathbf{w}}|\cdot] = \text{Var}[\bar{\mathbf{w}}|\cdot] \Phi^T (\text{Var}[\bar{\mathbf{z}}] + \Sigma_v)^{-1} \mathbf{y}$$

$$\text{Var}[\tilde{\mathbf{w}}|\cdot] = \left(\tilde{\lambda}^{-2} \mathbf{I}_n + \mathbf{D}_a^T (\text{Var}[\bar{\mathbf{z}}] + \Sigma_v)^{-1} \mathbf{D}_a\right)^{-1}$$

$$E[\tilde{\mathbf{w}}|\cdot] = \text{Var}[\tilde{\mathbf{w}}|\cdot] \mathbf{D}_a^T (\text{Var}[\bar{\mathbf{z}}] + \Sigma_v)^{-1} \mathbf{y}$$

D. The algorithm

In order to initialize the Gibbs sampler, the following steps are performed.

- Initialization of the parameters of the prior distributions for $\tilde{\lambda}^2$ and $\tilde{\lambda}^2$. Letting g_1, g_2, g_3 and g_4 be all equal to zero (implying that both the prior mean and variance are infinite) yields a non-informative prior.
- Initialization of the unobserved variables (i.e. $\bar{\mathbf{w}}, \tilde{\mathbf{w}}, 1/\tilde{\lambda}^2, 1/\tilde{\lambda}^2$) to values obtained by sampling the corresponding prior distribution or fixed either according to some a-priori knowledge or arbitrarily. In the present case, to speed up convergence, $\tilde{\lambda}^2$ and $\tilde{\lambda}^2$ are initialized with the so-called Empirical Bayes estimate, see [8], although an arbitrary initialization would work as well.

Subsequently, the following procedure has to be repeated iteratively so as to collect a sufficient amount of samples of both hyperparameters. During each iteration, the full conditional distributions are updated according to the values of the samples drawn during the preceding step. The structure of the general $i + 1$ -th iteration is described below.

- Compute of $\text{Var}[\bar{\mathbf{w}}|\cdot]$ and $E[\bar{\mathbf{w}}|\cdot]$
- Sample $\bar{\mathbf{w}}_{i+1}$ from the updated full conditional distribution of $\bar{\mathbf{w}}$
- Update of \bar{g}_1 and \bar{g}_2
- Sample $\tilde{\lambda}_{i+1}^2$ from the updated full conditional distribution of $\tilde{\lambda}^2$
- Compute of $\text{Var}[\tilde{\mathbf{w}}|\cdot]$ and $E[\tilde{\mathbf{w}}|\cdot]$
- Sample $\tilde{\mathbf{w}}_{i+1}$ from the updated full conditional distribution of $\tilde{\mathbf{w}}$
- Update of \bar{g}_3 and \bar{g}_4
- Sample $\tilde{\lambda}_{i+1}^2$ from the updated full conditional distribution of $\tilde{\lambda}^2$

After having extracted all samples, the initial ‘‘burn-in’’ part of the chain is discarded. The remaining h samples $\bar{\lambda}_i^2, \tilde{\lambda}_i^2, i = 1, \dots, h$, are used to compute ergodic averages $\bar{\lambda}_{MC}^2, \tilde{\lambda}_{MC}^2$ that approximate the posterior expectations of the hyperparameters.

IV. ESTIMATION OF THE CONTINUOUS-TIME SIGNALS

The MCMC algorithm described in the previous section hinges on the sampled model (3) and as such can estimate the typical and individual curves only in correspondence of the sampling knots. In this section, the posterior distribution of $\bar{\lambda}^2$ and $\tilde{\lambda}^2$ is exploited to obtain the posterior expectation and the posterior variance of the typical and individual curves at any time point.

A. The posterior expectation

The following results provide the point estimates for the typical curve and the individual ones.

Proposition 1: Let

$$\bar{c}_k^j = \frac{1}{h} \sum_{i=1}^h \bar{\lambda}_i^2 c_k^j(\bar{\lambda}_i^2, \tilde{\lambda}_i^2), \quad (4)$$

$$\begin{aligned} \mathbf{c}(\bar{\lambda}_i^2, \tilde{\lambda}_i^2) &:= [c_1^1 \dots c_{n_1}^1 \ c_1^2 \dots c_{n_2}^2 \dots c_1^N \dots c_{n_N}^N]^T := \\ &:= \text{Var}[\mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2]^{-1} \mathbf{y}, \quad i = 1, \dots, h, \end{aligned} \quad (5)$$

Then,

$$\hat{z}(t) := E[\bar{z}(t)|\mathbf{y}] \simeq \sum_{j=1}^N \sum_{k=1}^{n_j} \bar{\gamma}(t, t_k^j) \bar{c}_k^j \quad (6)$$

Proof: By the total probability theorem, the posterior expectation is

$$E[\bar{z}(t)|\mathbf{y}] = \iint E[\bar{z}(t)|\mathbf{y}, \bar{\lambda}^2, \tilde{\lambda}^2] p(\bar{\lambda}^2, \tilde{\lambda}^2|\mathbf{y}) d\bar{\lambda}^2 d\tilde{\lambda}^2$$

and, using the samples drawn by the MCMC algorithm, it can be approximated as

$$E[\bar{z}(t)|\mathbf{y}] \simeq \sum_{i=1}^h \frac{E[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2]}{h}$$

According to a well known formula for jointly Gaussian random variables,

$$\begin{aligned} E[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2] &= E[\bar{z}(t)|\bar{\lambda}_i^2, \tilde{\lambda}_i^2] + \\ &+ \text{Cov}[\bar{z}(t), \mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2] \text{Var}[\mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2]^{-1} (\mathbf{y} - E[\mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2]) \end{aligned}$$

Under the given assumptions, $E[\bar{z}(t)|\bar{\lambda}_i^2, \tilde{\lambda}_i^2] = 0$, $E[\mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2] = 0$, $\forall \bar{\lambda}_i^2, \forall \tilde{\lambda}_i^2$, while

$$\begin{aligned} \text{Cov}[\bar{z}(t), \mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2] &= \text{Cov}[\bar{z}(t), \bar{\mathbf{z}} + \tilde{\mathbf{z}} + \mathbf{v}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2] = \\ &= \text{Cov}[\bar{z}(t), \bar{\mathbf{z}}|\bar{\lambda}_i^2] = \bar{\lambda}_i^2 [\bar{\gamma}(t, t_1^1) \dots \bar{\gamma}(t, t_{n_N}^N)] \end{aligned}$$

$$\text{Var}[\mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2] = \text{Var}[\bar{\mathbf{z}}|\bar{\lambda}_i^2] + \text{Var}[\tilde{\mathbf{z}}|\tilde{\lambda}_i^2] + \Sigma_v$$

Then, recalling the definition of $\mathbf{c}(\bar{\lambda}_i^2, \tilde{\lambda}_i^2)$, we have that

$$E[\bar{z}(t)|\mathbf{y}] \simeq \sum_{i=1}^h \frac{1}{h} \sum_{j=1}^N \sum_{k=1}^{n_j} \bar{\gamma}(t, t_k^j) c_k^j(\bar{\lambda}_i^2, \tilde{\lambda}_i^2) \quad (7)$$

from which the thesis follows. \square

The computationally intensive step of the algorithm is (5), where the number of operations required to calculate $\text{Var}[\mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2]^{-1} \mathbf{y}$ scales as the cube of the number n of data. However, as shown in [11], $\text{Var}[\mathbf{y}]$ can be inverted in $O(n)$ operations, via Kalman filtering techniques, if \mathbf{y} is a vector obtained by sampling a process whose spectrum is rational. Analogous ideas can be used to efficiently evaluate the inverses needed in the MCMC simulation (Section III).

Proposition 2:

$$\hat{z}^j(t) := E[z^j(t)|\mathbf{y}] \simeq \hat{z}(t) + \sum_{k=1}^{n_j} \tilde{\gamma}(t, t_k^j) \tilde{c}_k^j$$

where

$$\tilde{c}_k^j = \frac{1}{h} \sum_{i=1}^h \tilde{\lambda}_i^2 c_k^j(\bar{\lambda}_i^2, \tilde{\lambda}_i^2) \quad (8)$$

and $c_k^j(\bar{\lambda}_i^2, \tilde{\lambda}_i^2)$ have been defined in Proposition 1.

Proof: All the considerations made about the estimation of the typical curve apply also to the estimation of the individual curves, with the only difference that $\bar{\gamma}(t, \tau)$ is replaced by $\tilde{\gamma}(t, \tau)$. \square

B. Population Regularization network

It is noteworthy that both $\hat{z}(t)$ and $\hat{z}^j(t)$ are obtained as linear combinations of autocovariance functions $\bar{\gamma}(t, \tau)$ and $\tilde{\gamma}(t, \tau)$, centered in the sampling knots $\tau = t_k^j$. This structure arises whenever Bayes estimation is used to reconstruct Gaussian processes [12], [13]. In particular, Poggio and Girosi (1990) coined the term *Regularization Network* (RN) to characterize estimators obtained via Gaussian processes estimation or Tychonov regularization (rather interestingly, the same final results are obtained by applying Bayesian estimation or reproducing kernel theory [14]). Also the estimators of Propositions 1 and 2 can be regarded as a kind of RN (see Fig. 1) in which there are two kinds of neurons. A first set of n neurons receive t as input and have $\bar{\gamma}(t, t_k^j)$ as activation functions. Their outputs are linearly combined through the weights \bar{c}_k^j to obtain $\hat{z}(t)$ as output. The individual shifts $\hat{z}^j(t)$ are estimated by a second set of n neurons having $\tilde{\gamma}(t, t_k^j)$ as activation functions and \tilde{c}_k^j as weights.

The weights \bar{c}_k^j and \tilde{c}_k^j are computed as ergodic averages, see (4), (8), of the coefficients $c_k^j(\bar{\lambda}_i^2, \tilde{\lambda}_i^2)$ which are obtained through the solution of a system of linear equations,

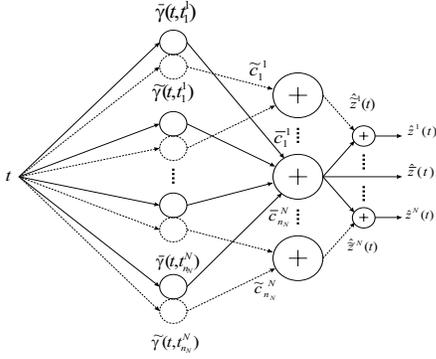


Fig. 1. Regularization Network (RN) with $2n$ neurons.

see (5), whose order is equal to the number of sampling knots. This is a major difference with respect to other networks such as MLP ones in which the weights are to be tuned via iterative methods. Recalling that the autocovariances $\tilde{\gamma}(t, t_k^j)$ are piecewise cubic polynomials, another important consequence of the regularization structure of the estimator is that the estimate $\hat{z}(t)$ of the typical curve is a cubic spline.

C. The posterior variance

In order to compute the posterior variance of $\bar{z}(t)$ recall that

$$\text{Var}[\bar{z}(t)|\mathbf{y}] = E[\bar{z}(t)^2|\mathbf{y}] - E[\bar{z}(t)|\mathbf{y}]^2 \quad (9)$$

Note that $E[\bar{z}(t)|\mathbf{y}]$ is known from (6). Moreover,

$$\begin{aligned} E[\bar{z}(t)^2|\mathbf{y}] &= \\ &= \int \bar{z}(t)^2 \iint p(\bar{z}(t)|\mathbf{y}, \bar{\lambda}^2, \tilde{\lambda}^2) p(\bar{\lambda}^2, \tilde{\lambda}^2|\mathbf{y}) d\bar{\lambda}^2 d\tilde{\lambda}^2 d\bar{z} = \\ &= \iint E[\bar{z}(t)^2|\mathbf{y}, \bar{\lambda}^2, \tilde{\lambda}^2] p(\bar{\lambda}^2, \tilde{\lambda}^2|\mathbf{y}) d\bar{\lambda}^2 d\tilde{\lambda}^2 \simeq \\ &\simeq \sum_{i=1}^h \frac{E[\bar{z}(t)^2|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2]}{h} \end{aligned}$$

where

$$\begin{aligned} E[\bar{z}(t)^2|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2] &= \text{Var}[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2] + \\ &+ E[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2]^2 \end{aligned}$$

The conditional expectation $E[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2]$ can be computed according to Proposition 1, while

$$\begin{aligned} \text{Var}[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2] &= \text{Var}[\bar{z}(t)|\bar{\lambda}_i^2] + \\ &- \text{Cov}[\bar{z}(t), \mathbf{y}|\bar{\lambda}_i^2] \text{Var}[\mathbf{y}|\bar{\lambda}_i^2, \tilde{\lambda}_i^2]^{-1} \text{Cov}[\bar{z}(t), \mathbf{y}|\bar{\lambda}_i^2]^T \end{aligned}$$

Analogous considerations hold for $\text{Var}[z^j(t)|\mathbf{y}]$.

Although this is more demanding from a computational point of view, the 5% and 95% percentiles of the posterior distributions of $\bar{z}(t)$ and $\hat{z}(t)$ could be calculated in a similar way. Indeed, by the total probability theorem, the posterior distribution of $\bar{z}(t)$ can be approximated as the average of h gaussian distributions whose mean and variance are $E[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2]$ and $\text{Var}[\bar{z}(t)|\mathbf{y}, \bar{\lambda}_i^2, \tilde{\lambda}_i^2]$, respectively.

V. EXAMPLE: ANALYSIS OF PHARMACOKINETIC DATA

The proposed population model was tested on a data set related to xenobiotics administration in 27 human subjects [7]. In the experiment, 8 samples were collected in each

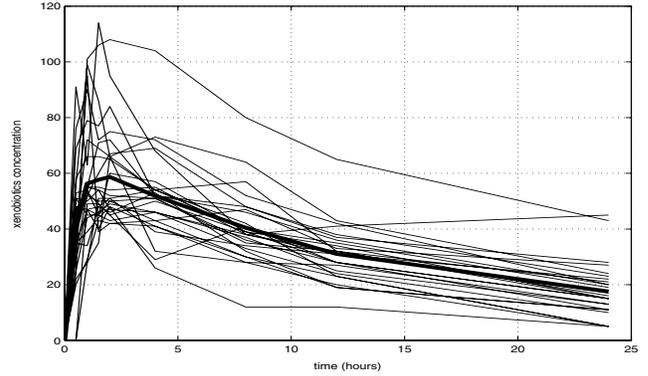


Fig. 2. Xenobiotics concentration data after a bolus in 27 human subjects: average curve (bold) and individual curves.

subject at $\{0.5, 1, 1.5, 2, 4, 8, 12, 24\}$ hours after a bolus administration. The data have a 10% coefficient of variation. To illustrate the population variability, the 27 experimental concentration curves are reported in Fig. 2, together with the average curve which, given the number of subjects, is a reasonable estimate of the typical curve $\bar{z}(t)$. Starting from these experimental data, different sampling schemes can be simulated by choosing proper subsets of the data. In particular, we adopted an example of a sparse and not well-designed sampling protocol: subject #2 is sampled at time points at $\{t_6, t_7, t_8\}$, #4 at $\{t_2, t_4, t_8\}$, #5 at $\{t_1, t_2\}$, #7 at $\{t_7\}$, #9 at $\{t_6\}$, #10 at $\{t_5\}$, #18 at $\{t_3, t_5\}$, #19 is fully sampled, #24 at $\{t_4, t_8\}$ and #26 at $\{t_1, t_3\}$ (25 samples in total).

The drug concentration is deterministically equal to zero at $t=0$, so that $\bar{\mathbf{X}}_0 = 0$, $\tilde{\mathbf{X}}_0 = 0$. Since the observed responses are not stationary but tend to be smoother towards the end of the experiment, the times were transformed logarithmically by defining a new time axis $t^{new} := \ln(t + 1)$. The analysis was carried out assuming that $a_1 = a_2 = -2$ (corresponding to an autocovariance $\text{Cov}[\tilde{z}^j(0), \tilde{z}^j(\tau)]$ of the shift process that is practically equal to zero for $\tau \geq 20$ hours), and estimating $\bar{\lambda}^2$ and $\tilde{\lambda}^2$ via MCMC with number of runs $h=1500$ ($\bar{\lambda}_{MC}^2 = 3553.3$, $\tilde{\lambda}_{MC}^2 = 16897$). In Fig. 3, the estimated typical curve with its $\pm 1.96\sqrt{\text{Var}[\bar{z}(t)|\mathbf{y}]}$ intervals is reported together with the data. In spite of the limited number of samples available to the algorithm (only

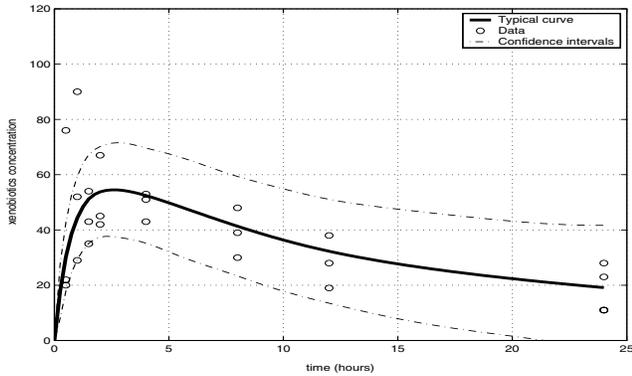


Fig. 3. Estimated typical curve (bold) with its 1.96 SD intervals.

25), the estimate of the typical curve is close to the population average curve showed in Fig. 2 that was based on the overall data set (216 observations). In Fig. 4, the estimate of the individual curve of subject #19 is shown together with its $\pm 1.96\sqrt{\text{Var}[z^{19}(t)|\mathbf{y}]}$ intervals. It is interesting to note the typical narrowing of the intervals in correspondence with the measured data. The other individual curves (not shown) are estimated with reasonable accuracy (taking into account the very limited number of data) but, as expected, the associated intervals are wider.

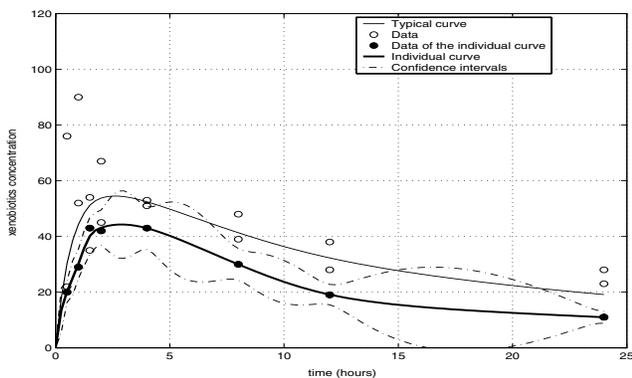


Fig. 4. Estimated individual curve (subject #19) with its 1.96 SD intervals. The estimated typical curve is also reported (dashed).

VI. CONCLUSIONS

An MCMC scheme that performs the estimation of a population of curves has been developed and tested on real pharmacokinetic data. To improve the computational efficiency, Monte Carlo sampling of the curves is performed

only at the sampling instants and the intersample posterior expectation and variance of the curves are subsequently reconstructed on the basis of the posterior distribution of the hyperparameters. It has been proven that, if the average curve is an integrated Wiener process with unknown variance, its estimate is a cubic spline whose knots coincide with the instants at which the individual curves are sampled.

VII. ACKNOWLEDGMENTS

This paper was partially supported by the PRIN Project “Metodi e algoritmi innovativi per l’identificazione e il controllo adattativo di sistemi tecnologici”.

REFERENCES

- [1] S. L. Beal and L. B. Sheiner, “Nonmem users guide,” University of California, San Francisco, 1998.
- [2] J. Wakefield, A. F. M. Smith, A. Racine-Poon, and A. Gelfand, “Bayesian analysis of linear and nonlinear population models using the gibbs sampler,” *Appl. Statist.*, vol. 41, pp. 201–221, 1994.
- [3] J. Wakefield and J. Bennett, “The Bayesian modelling of covariates for population pharmacokinetic models,” *JASA*, vol. 91, pp. 917–927, 1996.
- [4] R. Jelliffe, A. Schumitzky, M. Van Guilder, X. Wang, and R. Leary, “Population pharmacokinetic and dynamic models: parametric (p) and nonparametric (np) approaches,” in *14th IEEE Symposium on Computer-Based Medical Systems*, Bethesda (MD), 2001, pp. 407–412.
- [5] R. Leary, R. Jelliffe, A. Schumitzky, and M. Van Guilder, “An adaptive grid non-parametric approach to pharmacokinetic and dynamic (pk/pd) population systems,” in *14th IEEE Symposium on Computer-Based Medical Systems*, Bethesda (MD), 2001, pp. 389–394.
- [6] K. Park, D. Verotta, T. F. Blaschke, and L. B. Sheiner, “A semi-parametric method for describing noisy population pharmacokinetic data,” *J. Pharmacokin. Biopharm.*, vol. 25, pp. 615–642, 1997.
- [7] 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.*, vol. 29, no. 5/6, pp. 445–471, 2002.
- [8] M. Neve, G. De Nicolao, and L. Marchesi, “Nonparametric identification of pharmacokinetic population models via Gaussian processes,” *submitted*, 2004.
- [9] 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)*, 2003, pp. 53–55.
- [10] W. K. Hastings, “Monte Carlo sampling methods using Markov chain and their applications,” *Biometrika*, vol. 57, pp. 97–109, 1970.
- [11] G. De Nicolao and G. Ferrari-Trecate, “Regularization networks: Fast weight calculation via Kalman filtering,” *IEEE Transactions on Neural Networks*, vol. 12 (2), pp. 228–235, 2001.
- [12] T. Poggio and F. Girosi, “Networks for approximation and learning,” *Proc. IEEE*, vol. 78, pp. 1481–1497, 1990.
- [13] G. Wahba, *Spline Models for Observational Data*. Philadelphia: SIAM, 1990.
- [14] B. Ninness and H. Hjalmarsson, “On explicit characterisation of reproducing kernels with applications in estimation theory,” in *Proc. 16th International Symposium on Mathematical Theory of Networks and Systems (MTNS 2004)*, Leuven, 2004.