NONPARAMETRIC IDENTIFICATION OF PHARMACOKINETIC POPULATION MODELS VIA GAUSSIAN PROCESSES

M. Neve * **G. De Nicolao** * **L. Marchesi** *

* *Dipartimento di Informatica e Sistemistica, Università di Pavia, Italy*
< *marta.neve, giuseppe.denicolao* > *@unipv.it*

Abstract: Population models are used to describe the behaviour of different subjects belonging to a population and play an important role in drug pharmacokinetics. A nonparametric identification scheme is proposed in which both the average population response and the individual ones are modelled as Gaussian stochastic processes. Assuming that the average curve is an integrated Wiener process, it is shown that its estimate is a cubic spline. An Empirical Bayes (EB) algorithm for estimating both the typical and the individual curves is worked out. The model is tested on xenobiotics pharmacokinetic data.

Keywords: nonparametric identification, estimation theory, pharmacokinetic data, splines, neural networks, regularization.

1. INTRODUCTION

In biomedicine one is often faced with the problem of characterizing the average behaviour as well the inter-individual variability of a population of subjects. As an example, the analysis of population data is of primary importance in pharmacology, where drug responses collected in multiple subjects are used to obtain average and individual pharmacokinetic and pharmacodynamic models.

When it is possible to obtain a sufficient number of observations for each subject, model identification for each individual can be performed separately. However, in many cases there are technical, ethical and cost reasons that limit the number of samples that can be collected in each subject. Some examples are given by toxicokinetic studies as well as pharmacological experiments involving critical patients such as neonatal, pediatric or intensive care ones. If the individual models cannot be identified separately, it is necessary to resort to so-called "population methods" that provide the average and individual models from the joint analysis of all the available data.

Population methods can be divided into three main groups: parametric, semiparametric and nonparametric. In the parametric approach a structural model is assumed, e.g. a compartimental one, and the model parameters are regarded as random variables extracted from a distribution typical of the given population (Beal and Sheiner, 1998), (Wakefield *et al.*, 1994), (Wakefield and Bennett, 1996), (Jelliffe *et al.*, 2001), (Leary *et al.*, 2001) (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 other cases, for instance in the preliminary phases of a study, a structural model is not available and semiparametric or nonparametric techniques must be used. In the semiparametric approach, the response curves are modelled as regression splines (Park *et al.*, 1997), so that the

non-trivial problem of optimally placing the spline knots arises.

Recently, in order to develop a completely non-parametric approach, the individual curves have been modelled as discrete-time stochastic processes (e.g. random walks), reformulating the problem within the framework of Bayesian estimation (Magni *et al.*, 2002). This kind of model has also been used for the analysis of gene expression time series measured using DNA micro-arrays (Ferrazzi *et al.*, 2003). Since sampling schedules are usually not uniformly spaced in time, it would be more convenient to model the individual curves as continuous-time stochastic processes. In this paper we develop such a continuous-time population model. Model identification is carried out according to a so-called Empirical Bayes procedure. The method is tested on pharmacokinetic data related to xenobiotics administration in human subjects.

2. STOCHASTIC POPULATION MODEL

Consider the problem of estimating a family of scalar real-valued continuous-time functions $z^j(t)$, $j = 1, ..., N$, $t \geq 0$, on the basis of noisy samples taken at discrete instants. More precisely, assume that the following measurements are available

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

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 $E[v_k^j] = 0$, $Var[v_k^j] = (\sigma_k^j)^2$. In an experimental setting, the j-th curve $z^j(t)$ will be representative of the j-th subject (e.g. an impulse response obtained as a drug concentration profile in plasma after administration of a unit bolus). Note that the number and location of the sampling instants t_k^j may vary from subject to subject. In the following, each individual curve will be decomposed as

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

where $\bar{z}(t)$ is the "typical (average) curve" of the population and $\tilde{z}(t)$ is the "individual shift" with respect to the average behaviour. For ease of notation, the observations will be grouped in the vector

$$\mathbf{y} := [y_1^1 ... y_{n_1}^1\ y_1^2 ... y_{n_2}^2 ... y_1^N ... y_{n_N}^N]^T$$

Letting $n = n_1 + n_2 + ... + n_N$ be the total number of observations, $\mathbf{y}$ is an n-dimensional column vector. In a similar way it is possible to define the following vectors

$$\bar{\mathbf{z}} := [\bar{z}(t_1^1) ... \bar{z}(t_{n_1}^1)\ ... \bar{z}(t_1^N) ... \bar{z}(t_{n_N}^N)]^T$$

$$\tilde{\mathbf{z}} := [\tilde{z}^1(t_1^1) ... \tilde{z}^1(t_{n_1}^1)\ ... \tilde{z}^N(t_1^N) ... \tilde{z}^N(t_{n_N}^N)]^T$$

$$\mathbf{v} := [v_1^1 ... v_{n_1}^1\ v_1^2 ... v_{n_2}^2 ... v_1^N ... v_{n_N}^N]^T$$

Therefore, in vector notation, (1) can be rewritten as

$$\mathbf{y} = \bar{\mathbf{z}} + \tilde{\mathbf{z}} + \mathbf{v}$$

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

2.1 Typical and individual curves.

In the present paper a stochastic approach is adopted: the unknown functions are modelled as stochastic processes and the aim is to compute their posterior distributions given the observed data (note that the data are processed off-line, so that there is no need for the estimator to satisfy causality constraints).

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

In the following, $\bar{R}(t, \tau) := Cov[\bar{z}(t), \bar{z}(\tau)]$ and $\tilde{R}(t, \tau) := Cov[\tilde{z}^j(t), \tilde{z}^j(\tau)]$, $\forall j$, will denote the auto-covariance functions of the typical curve and the individual shifts, respectively. Recalling that $\tilde{z}^j(t)$ is a shift with respect to the typical response, it is reasonable to assume that $E[\tilde{z}^j(t)] = 0, \forall t, \forall j$. As for $\bar{z}(t)$, by properly scaling the data, it can be assumed without loss of generality that $E[\bar{z}(t)] = 0$. Since all the involved processes are jointly Gaussian, the posterior distributions are Gaussian as well. The following results provide the point estimates and the confidence intervals for the typical curve and the individual ones ($Var[\mathbf{y}]$ is the covariance matrix of the random vector $\mathbf{y}$).

Proposition 1.

$$\hat{\bar{z}}(t) := E[\bar{z}(t)|\mathbf{y}] = \sum_{j=1}^{N} \sum_{k=1}^{n_j} c_k^j \bar{R}(t, t_k^j) \qquad (2)$$

$$\hat{z}^j(t) := E[z^j(t)|\mathbf{y}] = \hat{\bar{z}}(t) + \sum_{k=1}^{n_j} c_k^j \tilde{R}(t, t_k^j) \qquad (3)$$

$$\mathbf{c} = \mathbf{\Sigma}_y^{-1} \mathbf{y} \qquad (4)$$

$$\mathbf{c} = [c_1^1 c_2^1 ... c_{n_1}^1 ... c_1^N ... c_{n_N}^N]^T$$

$$\mathbf{\Sigma}_y := Var[\mathbf{y}] = Var[\bar{\mathbf{z}}] + Var[\tilde{\mathbf{z}}] + \mathbf{\Sigma}_v$$

$$Var\,[\bar{\mathbf{z}}] = \begin{bmatrix} \bar{R}(t_1^1, t_1^1) & ... & \bar{R}(t_1^1, t_{n_N}^N) \\ ... & ... & ... \\ \bar{R}(t_{n_N}^N, t_1^1) & ... & \bar{R}(t_{n_N}^N, t_{n_N}^N) \end{bmatrix}$$

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

$$\tilde{\mathbf{R}}^j := \begin{bmatrix} \tilde{R}(t_1^j, t_1^j) & ... & \tilde{R}(t_1^j, t_{n_j}^j) \\ ... & ... & ... \\ \tilde{R}(t_{nj}^j, t_1^j) & ... & \tilde{R}(t_{n_j}^j, t_{n_j}^j) \end{bmatrix}$$

Proof: According to a well-known formula for jointly Gaussian random variables

$$E[\bar{z}(t)|\mathbf{y}] = E[\bar{z}(t)] + Cov[\bar{z}(t), \mathbf{y}]Var[\mathbf{y}]^{-1}(\mathbf{y} - E[\mathbf{y}])$$

Under the given assumptions, $E[\bar{z}(t)] = 0$, $E[\mathbf{y}] = 0$ and

$$Cov[\bar{z}(t), \mathbf{y}] = Cov[\bar{z}(t), \bar{\mathbf{z}} + \tilde{\mathbf{z}} + \mathbf{v}] = = Cov[\bar{z}(t), \bar{\mathbf{z}}] = [\bar{R}(t, t_1^1)...\bar{R}(t, t_{n_N}^N)]$$

The expressions for $\mathbf{\Sigma}_y$, $Var[\bar{\mathbf{z}}]$ and $Var[\tilde{\mathbf{z}}]$ are straightforwardly derived from the assumptions.□

Proposition 2.

$$Var[\bar{z}(t)|\mathbf{y}] = \bar{R}(t,t) - \bar{\mathbf{r}}\mathbf{\Sigma}_y^{-1}\bar{\mathbf{r}}^T$$

$$\bar{\mathbf{r}} := [\bar{R}(t, t_1^1)...\bar{R}(t, t_{n_N}^N)]$$

$$Var[z^j(t)|\mathbf{y}] = \bar{R}(t,t) + \tilde{R}^j(t,t) - (\bar{\mathbf{r}} + \tilde{\mathbf{r}}^j)\mathbf{\Sigma}_y^{-1}(\bar{\mathbf{r}} + \tilde{\mathbf{r}}^j)^T$$

$$\tilde{\mathbf{r}}^j := Cov[\tilde{z}^j(t), \tilde{\mathbf{z}}]$$

Proof: By a well-known formula

$$Var[\bar{z}(t)|\mathbf{y}] = Var[\bar{z}(t)] - Cov[\bar{z}(t), \mathbf{y}]Var[\mathbf{y}]^{-1}Cov[\bar{z}(t), \mathbf{y}]^T$$

Recalling that $\mathbf{y} = \bar{\mathbf{z}} + \tilde{\mathbf{z}} + \mathbf{v}$ and in view of the independency assumptions, the expression for $Var[\bar{z}(t)|\mathbf{y}]$ immediately follows. Analogous considerations hold for $Var[z^j(t)|\mathbf{y}]$.

2.2 Regularization Network interpretation.

It is interesting to note from (2) and (3) that the estimates $\hat{\bar{z}}(t)$ and $\hat{z}^j(t)$ are obtained as linear combinations of auto-covariance functions centered at the sampling knots t_k^j. This is the typical structure that comes out in the Bayesian estimation of Gaussian processes (Wahba, 1990), (Poggio and Girosi, 1990), (Girosi *et al.*, 1995), (Williams and Rasmussen, 1996). Remarkably, the same estimator can also be obtained via Tychonov regularization theory (Poggio and Girosi, 1990), (Girosi *et al.*, 1995). This explains why Poggio and Girosi (1990) have introduced the term *Regularization Network* (RN) to denote such estimators. Also the estimators of Proposition 1 can be regarded as RN's, although of a special type. Having to do with the identification of a population model, the number of neurons is $2n$ instead of n as in the standard RN, see Fig. 1. A first set of n neurons

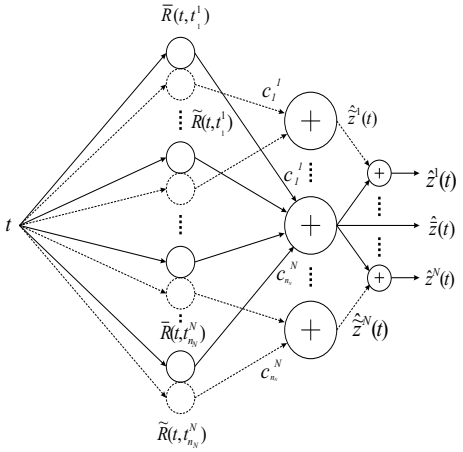

Fig. 1. Regularization Network structure of the estimator.

receive t as input and have $\bar{R}(t, t_k^j)$ as activation function. The estimate $\hat{\bar{z}}(t)$ of the typical curve is obtained by linearly combining these outputs through the weights c_k^j. A second set of n neurons, having $\tilde{R}(t, t_k^j)$ as activation functions, produce outputs that, combined again through the weights c_k^j, yield the estimates of the individual shifts $\hat{\tilde{z}}^j(t)$. The weight vector $\mathbf{c}$ is obtained as the solution of a system of n linear equations, see (4). This is an advantage with respect to other kinds of networks, such as Multi Layer Perceptrons, in which the weights have to be computed using iterative nonlinear optimization.

3. POPULATION SPLINES AND HYPER-PARAMETERS ESTIMATION

For the results of the previous section to be of practical use it is necessary to specify the statistics of the stochastic processes $\bar{z}(t)$, $\tilde{z}^j(t)$. If frequently sampled observations were available, the signal model could be identified by standard black-box parametric identification methods. On the other hand, population studies are often characterized by the scarcity of samples per subject. Therefore, it is necessary to introduce signal models that reflect the available a-priori knowledge.

3.1 Modelling the typical curve.

If it is only known that a signal is "smooth", it is a common practice to model it as an integrated Wiener process as done below.

Assumption 2.

$$\dot{\bar{x}}(t) = \bar{\mathbf{A}}\bar{x}(t) + \bar{\mathbf{B}}\bar{w}(t)$$

$$\bar{z}(t) = \bar{\mathbf{C}}\bar{x}(t)$$

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

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

The above model can describe signals whose initial conditions are deterministically known by setting $\bar{\mathbf{X}}_0 = 0$. Conversely, if the initial conditions are completely unknown, it suffices to let $\bar{\mathbf{X}}_0^{-1} = 0$ (although this requires some modification of the computational algorithms). The value of $\bar{\lambda}^2$ affects the regularity of the realizations so that smaller values correspond to smoother signals. The a-priori knowledge is seldom sufficient to specify $\bar{\lambda}^2$ so that it must be regarded as a "hyper-parameter" that will have to be estimated from the data.

Theorem 1. Under Assumption 2, $\hat{\bar{z}}(t)$ defined in Proposition 1 is a cubic spline with knots located in the sampling instants $\{t_1^1 t_2^1 ... t_{n_N}^N\}$.

Proof: It is well known that $\bar{\mathbf{X}}(t) = Var[\bar{x}(t)]$ is the solution of the differential Lyapunov equation

$$\dot{\bar{\mathbf{X}}}(t) = \bar{\mathbf{A}}\bar{\mathbf{X}}(t) + \bar{\mathbf{X}}(t)\bar{\mathbf{A}}^T + \bar{\lambda}^2\bar{\mathbf{B}}\bar{\mathbf{B}}^T$$
$$\bar{\mathbf{X}}(0) = \bar{\mathbf{X}}_0$$

Moreover,

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

In view of the definition of $\bar{\mathbf{A}}$, $\bar{\mathbf{B}}$, $\bar{\mathbf{C}}$, it follows that $\bar{R}(t,\tau)$, seen as a function of t, is a piecewise cubic polynomial. In particular, $\bar{R}(t,\tau)$ is continuous with all its derivatives everywhere but in $t = \tau$ where it is continuous up to the second derivative. Recalling that $\hat{\bar{z}}(t)$ in (2) is a linear combination of the functions $\bar{R}(t, t_k^j)$ (Proposition 1), the thesis immediately follows.□

In the literature, it is known that the conditional expectation of an integrated Wiener process given discrete observations is a cubic smoothing spline (Wahba, 1990). In some sense, Theorem 1 generalizes such a result to the analysis of population of signals so that it is reasonable to define $\hat{\bar{z}}(t)$ as a *population smoothing spline.*

3.2 Modelling the individual curves.

Coming now to the model for the individual shifts $\tilde{z}^j(t)$, the following assumption is in order.

Assumption 3. For $j = 1, ..., N$,

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

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

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

The statistics of $\tilde{z}(t)$ will depend 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 auto-covariance of $\tilde{z}^j(t)$. A possible drawback may be the difficulty in estimating two more hyper-parameters from the data. In this respect, it would be tempting to use a simpler model and describe also the individual shifts as integrated Wiener processes ($a_1 = 0$, $a_2 = 0$). However, observe that the measurements can be rewritten as

$$y_k^j = \bar{z}(t_k) + \bar{v}_k^j$$

where $\bar{v}_k^j := \tilde{z}^j(t_k) + v_k^j$. In other words, as far as the estimation of $\bar{z}(t)$ is concerned, $\bar{v}_k^j$ acts as measurement noise. If $\tilde{z}^j(t)$ were an integrated Wiener process, its variance would tend to infinity with t, and the confidence intervals for $\bar{z}(t)$ would diverge. For this reason, it is more convenient to select $\tilde{\mathbf{A}}$ with stable eigenvalues.

In view of Assumption 3, the calculation of $\tilde{R}(t,\tau)$ is completely analogous to that of $\bar{R}(t,\tau)$ described in the proof of Thm. 1.

3.3 Estimating the hyper-parameters.

When one is faced with a Bayesian estimation problem involving unknown hyper-parameters, a simple, yet effective, approach is to resort to the so-called Empirical Bayes method, see e.g. (MacKay, 1992). In the first step, a maximum likelihood estimate of the hyper-parameters is computed. Then, the Bayes estimate is calculated as if the hyper-parameters were deterministically known and equal to their maximum likelihood estimates. In the problem at hand, this leads to the following estimation algorithm, where $\theta = [\bar{\lambda}^2, \tilde{\lambda}^2, a_1, a_2]$ denotes the hyper-parameters vector.

Algorithm:

1. Let

$$\theta_{ML} := \arg\min_{\theta} \left\{\ln(\det(\mathbf{\Sigma}_y)) + \mathbf{y}^T\mathbf{\Sigma}_y^{-1}\mathbf{y}\right\}$$

2. Let

$$[\bar{\lambda}^2, \tilde{\lambda}^2, a_1, a_2]^T = \theta_{ML}$$

and compute $\hat{\bar{z}}(t)$ and $\hat{z}^j(t)$, $j = 1, ..., N$ according to Proposition 1.□

When the number of data per subject is scarce, identificability problems may be encountered in the estimation of the hyper-parameters. In particular it may be difficult to reliably estimate both a_1 and a_2 so that it may be convenient to impose some further constraint, e.g. a_1=a_2.

4. ANALYSIS OF PHARMACOKINETIC DATA

The proposed population model was tested on a data set related to xenobiotics administration in 27 human subjects. In the experiment, 8 samples were collected in each subject at {0.5, 1, 1.5, 2, 4, 8, 12, 24} hours after a bolus administration (Magni *et al.*, 2002). 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. 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 $\{t_6, t_7, t_8\}$, #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).

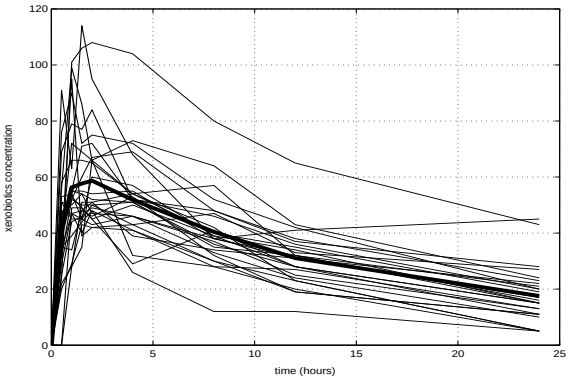

Fig. 2. Xenobiotics concentration data after a bolus in 27 human subjects: average curve (bold) and individual curves.

At time zero the drug concentration is known to be equal to zero, so that $\bar{\mathbf{X}}_0 = 0$, $\tilde{\mathbf{X}}_0 = 0$. In order to take into account the specific features of the data, some preprocessing was carried out. 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 $a_1 = a_2 = -2$. The hyper-parameters $\bar{\lambda}^2$ and $\tilde{\lambda}^2$ were estimated via maximum likelihood ($\bar{\lambda}^2_{ML} = 1785.8$, $\tilde{\lambda}^2_{ML} = 9257.2$). In Fig. 3, the estimated typical curve with its 95% confidence intervals is reported together with the data. Compared to the (piecewise linear) average curve computed on the basis of all the 216 observations (Fig. 2), the estimated typical curve (Fig. 3) appears to be a very satisfactory estimate. It has the advantage of being continuous up to the second derivative and, moreover, such a result was obtained using only 25 observations. In Fig. 4 the estimate of the individual curve of subject #19 is shown together with its confidence intervals. Rather interestingly, the confidence limits are narrower in correspondence of the sampling knots and become larger where less information is available. For the other individuals, reasonable estimates are obtained, although with larger confidence intervals (as expected since most subjects are very scarcely sampled).

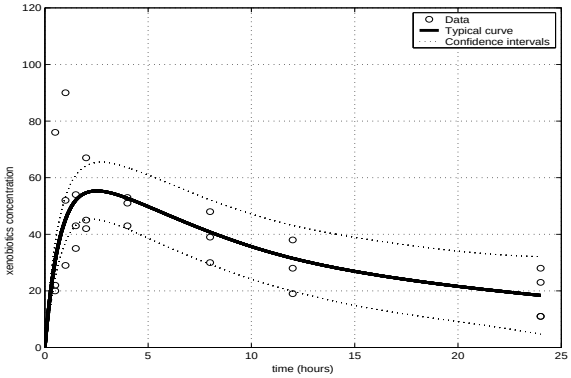

Fig. 3. Estimated typical curve (bold) with its 95% confidence intervals.

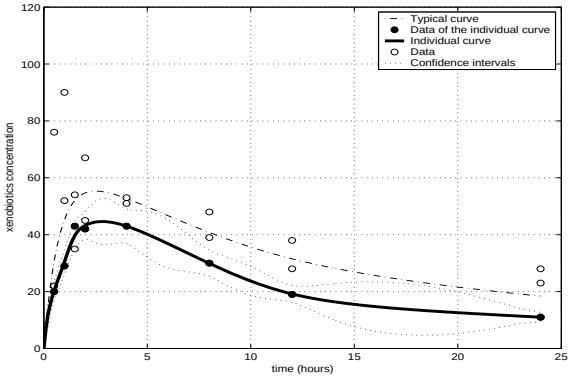

Fig. 4. Estimated individual curve of subject #19 (bold) with its 95% confidence intervals. The estimated typical curve is also reported (dash-dot).

5. CONCLUDING REMARKS

A new nonparametric continuous-time model for the population analysis of multiple experiments has been proposed. The typical curve as well as the individual ones are modelled as continuous-time Gaussian processes. If the statistics of the

processes are known, the posterior expectation given the data (the Bayes estimate), is obtained as the output of a regularization network, i.e. as the linear combination of auto-covariance functions centered at the sampling knots. The network weights are computed by solving a system of linear equations. Moreover, if the typical curve is modelled as an integrated Wiener process, its estimate is a cubic spline. In general, the statistics of the processes are not completely known and depend on some unknown hyper-parameters. Therefore, an Empirical Bayes scheme has been proposed: first the hyper-parameters are estimated via Maximum Likelihood (ML) and subsequently their ML estimates are plugged into the regularization network.

A first direction of future research will focus on the implementation of computationally efficient algorithms. In fact, the proposed scheme requires the solution of a system of linear equations and its computational complexity scales with the cube of the number of observations. By exploiting the state-space model it may be possible to work out algorithms based on Kalman filtering whose complexity scales linearly with the number of data. A second topic that is being currently investigated is the development of a truly Bayesian estimation procedure in which the hyper-parameters and the curves are estimated jointly using Markov Chain Monte Carlo algorithms.

6. ACKNOWLEDGMENTS

This paper was partially supported by the PRIN Project "Tecniche innovative per l'identificazione e il controllo adattativo di sistemi industriali".

REFERENCES

Beal, S. L. and L. B. Sheiner (1998). Nonmem users guide.

Ferrazzi, F., P. Magni and R. Bellazzi (2003). 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)*. pp. 53–55.

Girosi, F., M. Jones and T. Poggio (1995). Regularization theory and neural networks architectures. *Neural Computation* **7**, 219–269.

Jelliffe, R., A. Schumitzky, M. Van Guilder, X. Wang and R. Leary (2001). Population pharmacokinetic and dynamic models: parametric (p) and nonparametric (np) approaches. In: *14th IEEE Symposium on Computer-Based Medical Systems*. Bethesda (MD). pp. 407–412.

Leary, R., R. Jelliffe, A. Schumitzky and M. Van Guilder (2001). An adaptive grid nonparametric approach to pharmacokinetic and dynamic (pk/pd) population models. In: *14th IEEE Symposium on Computer-Based Medical Systems*. Bethesda (MD). pp. 389–394.

MacKay, D. J. C. (1992). Bayesian interpolation. *Neural Computation* **4**, 415–447.

Magni, P., R. Bellazzi, G. De Nicolao, I. Poggesi and M. Rocchetti (2002). Nonparametric AUC estimation in population studies with incomplete sampling: a Bayesian approach. *J. Pharmacokin. Pharmacodyn.* **29**(5/6), 445–471.

Park, K., D. Verotta, T. F. Blaschke and L. B. Sheiner (1997). A semiparametric method for describing noisy population pharmacokinetic data. *J. Pharmacokin. Biopharm.* **25**, 615–642.

Poggio, T. and F. Girosi (1990). Networks for approximation and learning. *Proc. IEEE* **78**, 1481–1497.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM. Philadelphia.

Wakefield, J., A. F. M. Smith, A. Racine-Poon and A. Gelfand (1994). Bayesian analysis of linear and nonlinear population models using the gibbs sampler. *Appl. Statist.* **41**, 201–221.

Wakefield, J. and J. Bennett (1996). The Bayesian modelling of covariates for population pharmacokinetic models. *JASA* **91**, 917–927.

Williams, C. K. J. and C. E. Rasmussen (1996). Gaussian processes for regression. In: *Advances in Neural Information Processing Systems* (D. S. Touretzky, M. C. Mozer and M. E. Hasselmo, Eds.). Vol. 8. Cambridge, MA: MIT Press.