Biometrical Journal **xxx** (2015) xxx, zzz–zzz / DOI: 10.1002/bimj.200100000

Inference in dynamic systems using B-splines and quasilinearized ODE penalties

Gianluca Frasso [*,1], **Jonathan Jaeger** [2], and **Philippe Lambert** [1,3]

[1] Institut des Sciences Humaines et Sociales, Méthodes Quantitatives en Sciences Sociales, Université de Liège *(Boulevard du Rectorat 7; B-4000 Liège; Belgium)*

[2] Nestlé Research Center, Lausanne, Switzerland

[3] Institut de Statistique, Biostatistique et Sciences Actuarielles, Université Catholique de Louvain, Belgium

Received zzz, revised zzz, accepted zzz

Nonlinear (systems of) ordinary differential equations (ODEs) are common tools in the analysis of complex one-dimensional dynamic systems. We propose a smoothing approach regularized by a quasilinearized ODE-based penalty. Within the quasilinearized spline based framework, the estimation reduces to a conditionally linear problem for the optimization of the spline coefficients. Furthermore, standard ODE compliance parameter(s) selection criteria are applicable. We evaluate the performances of the proposed strategy through simulated and real data examples. Simulation studies suggest that the proposed procedure ensures more accurate estimates than standard nonlinear least squares approaches when the state (initial and/or boundary) conditions are not known.

Key words: Nonlinear ordinary differential equations; Penalized splines; Quasilinearization.

1 Introduction

The analysis of one dimensional dynamic systems (e.g. systems evolving over time) represents a crucial task in many scientific fields. This kind of problem is usually tackled invoking differential calculus tools by defining (a system of) ordinary differential equation(s) which (analytic or numerical) solution(s) appropriately describes the observed dynamics.

In many circumstances, the parameters and the state (initial and/or boundary) conditions involved in the differential model synthesizing the observed phenomenon are not known. In these cases statistical procedures are applied to estimate them from noisy data. In general it is assumed that changes in the states $\boldsymbol{x}(t) \in \mathbb{R}^d$ of a dynamic system are governed by a set of differential equations:

$$\frac{d\boldsymbol{x}}{dt}(t) = \boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{\theta}), t \in [0; T], \tag{1}$$

where $\boldsymbol{f}$ is a known nonlinear function in the states $\boldsymbol{x}$ and $\boldsymbol{\theta} \in \mathbb{R}^q$ an unknown vector of parameters. It is assumed that only a subset $\mathcal{J} \subset \{1, \ldots, d\}$ of the d state functions $\boldsymbol{x}$ is observed at time point t_{jk}, $j \in \mathcal{J}$, $k = 1, \ldots, n_j$ with additive measurement error ϵ_{jk}. We denote by $y_{jk} = x_j(t_{jk}) + \tau_j^{-1/2}\epsilon_{jk}$ the corresponding measurement. For simplicity, we assume that the error terms ϵ_{jk} are distributed according to a standardized Gaussian distribution. Our goal is to estimate the ODE parameters $\boldsymbol{\theta}$, the precision of measurements $\boldsymbol{\tau} = \text{vec}(\tau_j, j \in \mathcal{J})$ and to approximate, using basis expansion, the solution of the ODE model given in Eq. (1) from the (noisy) observations.

Many approaches have been proposed in the literature to estimate these sets of unknowns. The first reference in this field is probably Hotelling (1927) that tackled this task from a regression point of view. Following this idea, Biegler et al. (1986) proposed a nonlinear least squares approach for the estimation of

*Corresponding author: e-mail: gianluca.frasso@ulg.ac.be

the ODE parameters using suitable numerical schemes to approximate the state function(s). This approach suffers of some drawbacks. It can be computationally demanding to deal with complex differential systems and the quality of the final estimates can be severely influenced by possible error propagation effects (related to the adopted numerical scheme and to the approximation of the state, initial and/or boundary, conditions). Finally these techniques become unpractical for inferential purposes in complex dynamics.
For these reasons the penalized smoothing collocation approach of Ramsay et al. (2007) represents an attractive alternative. This estimation procedure consists in a compromise between a flexible (regression based) description of the recorded measurements and the (numerical) solution of the hypothesized differential system obtained through a collocation scheme. In particular, this framework exploits a B-spline approximation of the state function. The flexibility induced by the B-spline approximation is then counterbalanced by a penalty related to the differential model. The ODE-penalty term is defined as the integral of the ODE operator evaluated at the B-spline approximation of the state function. It measures the fidelity of the B-spline approximation to the ODE model. An ODE-compliance parameter balances between the fidelity to the data and the compliance of the B-spline approximation to the ODE model. This approach can be adopted both in case of linear and nonlinear differential equations and does not require the estimation of the state conditions.
In contrast to the linear case, estimation in nonlinear systems using the approach by Ramsay et al. (2007) requires a nonlinear least squares step and the application of the implicit function theorem to obtain point and interval estimates of the ODE parameters. Furthermore, even if the compliance parameter still tunes the balance between data fitting and ODE model fidelity, the application of standard selection procedures can be unpractical dealing with nonlinear ODE systems.
In this paper, we overcome these challenges by introducing a quasilinearization scheme (Bellman and Kalaba, 1965) in the estimation procedure. Our quasilinearized ODE-P-spline (QL-ODE-P-spline) approach can be viewed as a convenient generalization of the P-spline collocation procedure of Ramsay et al. (2007) as it permits an explicit link between the spline coefficients and the unknown differential parameters. Furthermore, standard approaches for selecting/optimizing the ODE-compliance parameters become directly applicable.
In this paper we consider two differential problems as working examples to illustrate our proposal and evaluate its performances and robustness through simulations: 1) a simple first order ODE having a closed form solution; 2) the well known Van der Pol system of ODEs. As first example we consider the following *first order initial value problem*

$$\begin{cases} \dfrac{dx}{dt}(t) &= (\theta_1 - 2\theta_2 t)\, x^2(t) \\ x(0) &= x_0, \end{cases} \tag{2}$$

with explicit solution

$$x(t) = \frac{1}{\theta_2 t^2 - \theta_1 t + x_0^{-1}}. \tag{3}$$

The second working example is represented by the *Van der Pol system*:

$$\begin{cases} \dfrac{dx_1}{dt}(t) &= \theta\left(x_1(t) - \dfrac{1}{3}x_1^3(t) - x_2(t)\right) \\ \dfrac{dx_2}{dt}(t) &= \dfrac{1}{\theta}x_1(t), \\ \boldsymbol{x}(t_0) &= \boldsymbol{s}_{t_0} \text{ for } t_0 \in \mathcal{T}_0, \end{cases} \tag{4}$$

where $\mathcal{T}_0$ denotes the subset of observed times where the state conditions are imposed. The state function $x_1(t)$ describes a non-conservative oscillator with nonlinear damping. It was proposed to describe the dynamics of current charge in a nonlinear electronic oscillator circuit (Van der Pol, 1926). This model also

found applications in other fields such as biology and seismology. No exact solution can be derived when the parameter θ differs from zero.
The paper is organized as follows. In Section 2 the key features of the generalized profiling estimation procedure proposed by Ramsay et al. (2007) are briefly recalled. In Section 3, the introduction of a quasilinearization step in the profiling framework is illustrated to highlight the strengths of our QL-ODE-P-spline method. Its properties are studied using simulations in Section 4. We conclude the paper with the analysis of two real(istic) datasets in Section 5 followed by a discussion in Section 6.

2 Generalized profiling estimation in a nutshell

In this section, the key steps of the profiling procedure proposed by Ramsay et al. (2007) are briefly reminded. First of all, each state function involved in the system of (nonlinear) differential equations is approximated using a B-spline basis function expansion:

$$\widetilde{x_j}\left(t|\boldsymbol{\alpha}\right) = \sum_{k=1}^{K_j} b_{jk}\left(t\right)\alpha_{jk} = \left[\boldsymbol{b}_j\left(t\right)\right]^\top \boldsymbol{\alpha}_j$$

where $\boldsymbol{\alpha}_j$ is a K_j-vector of spline coefficients and $\boldsymbol{b}_j\left(t\right)$ is the K_j-vector of B-spline basis functions evaluated at time t. The variation of the B-spline approximations are shrunk by introducing an ODE-based penalty. More precisely, for each differential equation of the system, a penalty term is introduced:

$$PEN_j\left(\boldsymbol{\alpha}|\boldsymbol{\theta}\right) = \int \left(\frac{d\widetilde{x}_j}{dt}\left(t\right) - f_j\left(t, \widetilde{\boldsymbol{x}}, \boldsymbol{\theta}\right)\right)^2 dt, \tag{5}$$

where $\boldsymbol{\alpha} = \text{vec}\left(\boldsymbol{\alpha}_j; j = 1, \ldots, d\right)$. The full fidelity measure to the ODE-model is then defined as:

$$PEN\left(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\gamma}\right) = \sum_{j=1}^{d} \gamma_j PEN_j\left(\boldsymbol{\alpha}|\boldsymbol{\theta}\right). \tag{6}$$

The compromise between data fitting and model fidelity is tuned by the ODE-compliance parameters $\boldsymbol{\gamma}$. For values of γ_j close to zero, the final estimates tend to overfit the data by satisfying nearly exclusively the goodness of fit criterion. For $\gamma_j \to \infty$, the final approximation tends to mimic the solution of the ODE model used to define the penalty.
The ODE parameters and the spline coefficients are estimated by profiling the likelihood. This approach requires to optimize the choice of the ODE parameters by considering the spline coefficients as nuisance parameters. For given values of the ODE parameters $\boldsymbol{\theta}$, of the precisions of measurement $\{\tau\}_{j\in\mathcal{J}}$ and of the ODE-compliance parameters $\boldsymbol{\gamma}$, the optimal spline coefficients are found as the maximizer of:

$$J\left(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{y}\right) = \sum_{j\in\mathcal{J}} \left\{\frac{n_j}{2}\log\left(\tau_j\right) - \frac{\tau_j}{2}\left\|\boldsymbol{y}_j - \widetilde{x}_j\left(\boldsymbol{t}\right)\right\|^2\right\} - \frac{1}{2}PEN\left(\boldsymbol{\alpha}|\boldsymbol{\theta}, \boldsymbol{\gamma}\right). \tag{7}$$

In a second step, given the current value of the ODE compliance parameters, the vector $\boldsymbol{\theta}$ is estimated together with the precisions of measurement optimizing the following data fitting criterion:

$$H\left(\boldsymbol{\theta}, \boldsymbol{\tau}|\boldsymbol{y}, \widehat{\boldsymbol{\alpha}}\right) = \sum_{j\in\mathcal{J}} \left\{\frac{n_j}{2}\log\left(\tau_j\right) - \frac{\tau_j}{2}\left\|\boldsymbol{y}_j - \boldsymbol{B}_j\widehat{\boldsymbol{\alpha}}_j\left(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{y}\right)\right\|^2\right\}, \tag{8}$$

with $\boldsymbol{B}_j$ the B-spline matrix defined for the jth state function. Following Eilers and Marx (2010), we suggest to define the B-spline bases using a generous number of equidistant knots. For the degree of the B-splines we found that a useful rule of thumb is to set it equal to the degree of the differential equation

plus two. This can be justified by examining the properties of B-spline functions and their derivatives. The hth order derivative of a B-spline basis of degree q is obtained by finite differentiation of splines of $q-h$ degree. A qth degree B-spline is $q-1$ times continuously differentiable over the domain span (see e.g. Golub and Ortega, 1992; Dierckx, 1995). By defining bases of degree equal to the order of the ODE plus two, we ensure that the bases used to approximate the highest order derivative appearing in the dynamic system are at least one time continuously differentiable over the entire domain.
As the definition of the penalty term $PEN\left(\boldsymbol{\alpha}|\boldsymbol{\theta},\boldsymbol{\gamma}\right)$ is based on (a system of) nonlinear differential equation(s), the minimizer of $J\left(\boldsymbol{\alpha}|\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{y}\right)$ cannot be found analytically. This makes the dependence of H on $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ partially implicit. Therefore, the function H can only be optimized by combining a non linear least squares solver with the implicit function theorem. Similar arguments are involved in the approximation of the variance-covariance matrix of $\widehat{\boldsymbol{\theta}}$ and in the estimation of the model effective dimension (see Ramsay et al. (2007) for details). Finally, the selection of the optimal ODE-compliance parameters is a non trivial task as common smoothing parameter selection criteria cannot be applied.

3 Quasilinearized ODE-P-spline approach

When nonlinearity becomes an issue for the estimation process, it is convenient to approximate nonlinear functions with linear surrogates (obtained for example through first order Taylor expansion). Quasilinearization (Bellman and Kalaba, 1965) generalizes this idea to approximate the solution of differential equations. In the generalized profiling estimation procedure, provided that differential equations are linear, the estimation of $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ can be made by iterating simple least squares procedures with ODE compliance parameters selected using standard methods. From the discussion at the end of Section 2 it appears that dealing with nonlinear ODEs is more challenging. Our main suggestion is to amend the estimation procedure using quasilinearization.

3.1 Quasilinearization of the penalty

We propose to linearize the nonlinear part of the ODE model using an iterative approximation procedure. More precisely, given the set of spline coefficients $\boldsymbol{\alpha}^{(i)}$ at iteration i and the induced estimates of the j-th state function $\widetilde{x}_j^{(i)}$, we propose to approximate the j-th component of the ODE penalty term at iteration $i+1$ by:

$$\begin{aligned} PEN_j\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\alpha}^{(i)},\boldsymbol{\theta}\right) &\approx \int\left(\frac{d\widetilde{x}_j^{(i+1)}}{dt}(t) - f_j\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) - \sum_{k=1}^{d}\left(\widetilde{x}_k^{(i+1)}(t)\right.\right. \\ &- \left.\left.\widetilde{x}_k^{(i)}(t)\right)\frac{\partial f_j}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\right)^2 dt. \end{aligned} \tag{9}$$

Then, the overall fidelity measure to the ODE in Eq. (6) can be approximated by:

$$\begin{aligned} PEN\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\alpha}^{(i)},\boldsymbol{\theta},\boldsymbol{\gamma}\right) &= \sum_{j=1}^{d}\gamma_j PEN_j\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\alpha}^{(i)},\boldsymbol{\theta}\right) \\ &= \left(\boldsymbol{\alpha}^{(i+1)}\right)^\top \boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)\boldsymbol{\alpha}^{(i+1)} \\ &+ \left(\boldsymbol{\alpha}^{(i+1)}\right)^\top \boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right) + l\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right), \end{aligned} \tag{10}$$

where $\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ plays the role of a penalty matrix, $\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ is the penalty vector and $l\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ is a constant not depending on the current values of spline coefficients but only on the previous ones (see Appendix A for the explicit definition of $\boldsymbol{R},\boldsymbol{r}$ and l). In contrast to the inner optimization proposed in

Ramsay et al. (2007), the substitution of (10) in (7) makes the optimization of the criterion J with respect to the spline coefficients a standard least squares problem that does not require any sophisticated mathematical analysis tool. Indeed, Eq. (7) becomes a quadratic form in $\boldsymbol{\alpha}^{(i+1)}$ which optimization does not require the implicit function theorem anymore.
Similarly to Ramsay et al. (2007) and Frasso et al. (2013), the estimation process for the ODE parameters and the spline coefficients implies the iteration of two profiling steps. First, given the last available estimates $\boldsymbol{\alpha}^{(i)}$ and for given vectors of ODE parameters $\boldsymbol{\theta}^{(i)}$, precisions of measurements $\boldsymbol{\tau}^{(i)} = \text{vec}\left(\tau_j^{(i)}; j \in \mathcal{J}\right)$ and compliance parameters $\boldsymbol{\gamma}^{(i)}$, the spline coefficients $\boldsymbol{\alpha}^{(i+1)}$ are updated by maximizing the approximated J criterion:

$$\begin{aligned}\boldsymbol{\alpha}^{(i+1)} &= \underset{\boldsymbol{\alpha}}{\text{argmax}}\, J\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\theta}^{(i)}, \boldsymbol{\tau}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\alpha}^{(i)}, \boldsymbol{y}\right) \\ &= \left(\boldsymbol{G}^{(i)} + \boldsymbol{R}\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\alpha}^{(i)}\right)\right)^{-1}\left(\boldsymbol{g}^{(i)} - \boldsymbol{r}\left(\boldsymbol{\theta}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\alpha}^{(i)}\right)\right), \end{aligned} \tag{11}$$

where $\boldsymbol{G}^{(i)} = \text{diag}\left(\boldsymbol{Z}_j^{(i)}; j = 1, \ldots, d\right)$ is a block diagonal matrix with $\boldsymbol{Z}_j^{(i)} = \tau_j^{(i)} \boldsymbol{B}_j^\top \boldsymbol{B}_j$ if the j-th state function is observed and the null $K_j \times K_j$-matrix otherwise and $\boldsymbol{g}^{(i)} = \text{vec}\left(\boldsymbol{z}_j^{(i)}; j = 1, \ldots, d\right)$ with $\boldsymbol{z}_j^{(i)} = \tau_j^{(i)} \boldsymbol{B}_j^\top \boldsymbol{y}_j$ if the j-th state function is observed or null (K_j-vector) otherwise. Here, $\boldsymbol{y}_j$ is the n_j-vector of all observations for the j-the state function and $\boldsymbol{B}_j$ is a B-spline matrix of dimension $n_j \times K_j$.
The values of $\boldsymbol{\theta}$ and $\boldsymbol{\tau}$ are updated by optimizing H given $\boldsymbol{\alpha}^{(i+1)}$.
Thanks to the quasilinearization, the selection of the compliance parameters can be performed using standard methods such as AIC, BIC or (generalized) cross-validation. On the other hand, these parameters can be defined as the ratio of the estimated error variances and the current variance of the penalty (points (**c**) and (**d**) of Algorithm 1) and optimized using the EM-type algorithm proposed by Schall (1991). In this way we avoid the computation of a selection criterion over a grid of possible values and treat the ODE-parameter selection as an inner step of the estimation algorithm.
The QL-ODE-P-spline estimation procedure can be summarized as follows:
The following convergence criterion was used throughout to stop the iterative process of Algorithm 1:

$$\left[\max\left(\left|\frac{\boldsymbol{\alpha}^{(i+1)} - \boldsymbol{\alpha}^{(i)}}{\boldsymbol{\alpha}^{(i)}}\right|\right), \max\left(\left|\frac{\boldsymbol{\gamma}^{(i+1)} - \boldsymbol{\gamma}^{(i)}}{\boldsymbol{\gamma}^{(i)}}\right|\right), \max\left(\left|\frac{\boldsymbol{\theta}^{(i+1)} - \boldsymbol{\theta}^{(i)}}{\boldsymbol{\theta}^{(i)}}\right|\right), \max\left(\left|\frac{\boldsymbol{\tau}^{(i+1)} - \boldsymbol{\tau}^{(i)}}{\boldsymbol{\tau}^{(i)}}\right|\right)\right] < 10^{-4}. \tag{12}$$

In order to simplify the notation, in the rest of our discussion the iteration indexes for the $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}$ vectors will be omitted when not necessary.
The initial values for the ODE parameters and the precision of measurements can be either chosen by the analyst on the basis of prior knowledge or estimated from the raw measurements. In the latter case different alternatives are possible. One possibility is to invert the ODE system using the spline coefficients estimated by P-splines (with optimal smoothing parameter selected through standard methods such as generalized cross validation, AIC, BIC and others). Alternatively, the estimates of a NLS analysis (based on the numerical solution of the ODE) can be used. The ODE compliance parameters can be set at large values (as in Ramsay et al. (2007)) if the interest is exclusively focused on ODE parameter estimation. Then, step (**d**) in Algorithm 1 can be skipped.
The proposed procedure requires also an initial vector of spline coefficients: we recommend to use a standard P-spline smoother (Eilers and Marx, 1996) to start with reasonable initial coefficients $\boldsymbol{\alpha}^{(0)}$. The ODE model combined with the initial ODE parameters ($\boldsymbol{\theta}^{(0)}$) and the initial spline coefficients associated to the observed state functions can be used to deduce initial spline coefficients for the unobserved state functions. On the other hand, according to our experience, the quality of the final estimates is not dramatically influenced by the choice of $\boldsymbol{\alpha}^{(0)}$.

Data: Set of (noisy) measurements driven by a dynamic system.
Result: Estimation of optimal $\boldsymbol{\alpha}$, $\boldsymbol{\theta}$, $\boldsymbol{\gamma}$ and $\boldsymbol{\tau}$ via QL-ODE-P-splines.
Initialization: Smooth the data using standard P-splines to estimate the initial spline coefficients $\boldsymbol{\alpha}^{(0)}$. Set the initial $\boldsymbol{\theta}^{(0)}$, $\boldsymbol{\tau}^{(0)}$ and $\boldsymbol{\gamma}^{(0)}$.
while *Convergence is not achieved* **do**

(**a**) Given the current $\boldsymbol{\alpha}^{(i)}$ and $\boldsymbol{\gamma}^{(i)}$ linearize the ODE based penalties PEN_j (see Appendix A) ;

(**b**) Compute $\boldsymbol{\alpha}^{(i+1)}$ as Eq. (11) using the linearized penalties and compute $\widetilde{x}^{(i+1)}(\boldsymbol{t}) = \boldsymbol{B}\boldsymbol{\alpha}^{(i+1)}$;

(**c**) Update $\boldsymbol{\tau}^{(i+1)} = \dfrac{N - \mathbf{ED}^{(i+1)}}{\|\boldsymbol{y} - \widetilde{x}^{(i+1)}(\boldsymbol{t})\|^2}$ (the effective dimension is computed as $\mathbf{ED}^{(i+1)} = \text{tr}\left[\boldsymbol{B}\left(\boldsymbol{G}^{(i)} + \boldsymbol{R}(\boldsymbol{\theta}^{(i)}, \boldsymbol{\gamma}^{(i)}, \boldsymbol{\alpha}^{(i)})\right)^{-1}\boldsymbol{\tau}^{(i)}\boldsymbol{B}^{\top}\right]$ and $\boldsymbol{G}^{(i)} = \boldsymbol{\tau}^{(i)}\boldsymbol{B}^{\top}\boldsymbol{B}$ following Hastie and Tibshirani (1990)) ;

(**d**) Compute $\left[\boldsymbol{\sigma}^2_{PEN}\right]^{(i+1)} = \dfrac{PEN\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\alpha}^{(i)}, \boldsymbol{\theta}^{(i)}, \boldsymbol{\gamma}^{(i)}\right)}{\mathbf{ED}^{(i+1)}}$ and ODE compliance parameters $\boldsymbol{\gamma}^{(i+1)} = 1/[\boldsymbol{\sigma}^2_{PEN}]^{(i+1)}$;

(**e**) Given the current values of the other unknowns, update the ODE parameters by minimizing Eq. (8);

(**f**) Compute the convergence criterion in Eq. (12).

end

Algorithm 1: Quasilinearized ODE-P-spline estimation algorithm.

This issue is illustrated by the results of a simulation study summarized in Table 1. One hundred data sets have been generated by adding a zero mean Gaussian noise to the solutions of the ODE problems presented in Section 1. These data have been analyzed using QL-ODE-P-splines with initial ODE parameters set either to their true values $\boldsymbol{\theta}^*$, to $0.5\boldsymbol{\theta}^*$ or to $2\boldsymbol{\theta}^*$. The initial spline coefficients have been obtained: by smoothing of the raw measurements with standard P-splines (Strategy 1), or by assuming constant state functions (Strategy 2). The smoothers have been built on cubic B-splines defined over 40 equidistant internal knots to approximate the state function in the first order ODE model (Example 1) and using fourth order spline functions computed on 150 equidistant internal knots for the simulation of the Van der Pol system (Example 2). Samples of $N = 50$ and $N = 100$ observations have been generated for Example 1 and 2 respectively.

Figure 1 shows possible estimation steps for Example 1 starting with a constant state function and wrong initial ODE parameters. These plots clearly show the impact of the bad initial guesses (upper left panel). At the second estimation step, the ODE-compliance parameter γ becomes small giving more weight to the goodness-of-fit term in Eq. (7). Then, the (fitted) black curve tends to overfit the data with estimated ODE parameters different from the true ones ($\theta_1^* = \theta_2^* = 1$). As the number of iterations increases, the ODE compliance parameter also increases, yielding more weight to the ODE-based penalty in Eq. (7) and ensuring that the estimated state functions and ODE parameters get closer and closer to their true values. Figure 2 shows similar results for Example 2 (based on the Van der Pol system). It is remarkable that, even with initial spline coefficients chosen without taking into account the ODE model and/or the observations, we properly approximate both the observed ($x_1(t)$) and unobserved ($x_2(t)$) state functions.

Table 1 shows the (average) root mean squared error w.r.t. the data ($\overline{\text{RMSE}(\boldsymbol{y})} = 0.01\sum_{k=1}^{100}\sqrt{N^{-1}\|\boldsymbol{y}_k - \widetilde{\boldsymbol{x}}_k\|^2}$) and the RMSE w.r.t. the state function ($\overline{\text{RMSE}(\boldsymbol{x})} = 0.01\sum_{k=1}^{100}\sqrt{N^{-1}\|\boldsymbol{x}_k - \widetilde{\boldsymbol{x}}_k\|^2}$): the influence of the initial guesses on the goodness of the final fit appears negligible. Analogously, the compliance of the approximated state functions to the simulated ones does not seem affected by the specification of the initial

settings. On average, the choice of the initial values for the spline coefficients seems to have a limited impact on the number of iterations required to achieve convergence ($\overline{\#\,\text{Iter}}$). From these results, we can conclude that Algorithm 1 ensures estimates robust with respect to the choice of initial parameters (ODE-parameters and spline coefficients).
Finally notice that, like in the framework of Ramsay et al. (2007), our QL-ODE-P-spline procedure does not require the knowledge (or approximation) of initial state conditions. This is a major advantage over traditional numerical procedures such as Runge-Kutta based nonlinear least squares approaches. On the other hand, as pointed by Frasso et al. (2013), in some cases these conditions arise as characteristic of the data or as theoretical requirements and cannot be ignored in order to obtain meaningful estimates of the parameters and relevant approximations to the state function(s). In these cases the state conditions can be included in the framework above (see Appendix C).

[Figure 1 about here.]

[Figure 2 about here.]

[Table 1 about here.]

3.2 Approximate sampling variances of $\widehat{\theta}$ and $\widetilde{x}$ at convergence

In step (**e**) of Algorithm 1 the unknown ODE parameters are estimated by maximizing (8). Plugging-in the (current) precision parameter $\widehat{\tau}_j = \dfrac{(N - \mathbf{ED})}{\left\|\boldsymbol{y}_j - \boldsymbol{B}_j\widehat{\boldsymbol{\alpha}}_j\left(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{y}\right)\right\|^2}$ into $H(\boldsymbol{\theta}, \boldsymbol{\tau}|\boldsymbol{y}, \boldsymbol{\alpha})$, the criterion to be maximized becomes:

$$H\left(\boldsymbol{\theta}|\boldsymbol{y}, \widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\tau}}\right) = \sum_{j\in\mathcal{J}}\left\{-\frac{\widehat{\tau}_j}{2}\left\|\boldsymbol{y}_j - \boldsymbol{B}_j\widehat{\boldsymbol{\alpha}}_j\left(\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{y}\right)\right\|^2\right\}, \tag{13}$$

from which, at convergence of the QL-ODE-P-spline algorithm, (the inverse of) the Fisher information matrix (FIM) can be computed to approximate the variance of the estimated ODE parameters:

$$\text{var}(\widehat{\boldsymbol{\theta}}) \approx \left[-\text{E}\left(\frac{\partial^2 H}{\partial\boldsymbol{\theta}^2}\Big|_{\widehat{\boldsymbol{\theta}}}\right)\right]^{-1}. \tag{14}$$

This approximation is analogous to the one in Rodriguez-Fernandez et al. (2006) and Xue et al. (2010) but the quasilinearization of the penalty greatly simplifies the computation of the second derivatives in Eq. 14 (see Appendix B for details and for comparisons with the estimator suggested by Ramsay et al. (2007)).
The proposed QL-ODE-P-spline also simplifies the approximation of pointwise confidence bands for $\widetilde{\boldsymbol{x}}$. Indeed, by defining the hat matrix as $\boldsymbol{A} = \boldsymbol{B}\left(\boldsymbol{G}(\widehat{\boldsymbol{\tau}}) + \boldsymbol{R}(\boldsymbol{\theta}, \boldsymbol{\gamma}, \widehat{\boldsymbol{\alpha}}^{(i)})\right)^{-1}\widehat{\boldsymbol{\tau}}\boldsymbol{B}^\top$, one obtains $\text{var}(\widetilde{\boldsymbol{x}}) \approx \widehat{\boldsymbol{\tau}}^{-1}\left(\boldsymbol{A}^\top\boldsymbol{A}\right)$. Approximate pointwise confidence bands can be computed using $\widetilde{\boldsymbol{x}} \pm 2\sqrt{\text{diag}\left[\text{var}(\widetilde{\boldsymbol{x}})\right]}$.

4 Practical implementation and simulation

In this section the results of simulation studies based on the wo examples presented in Section 1 are reported to illustrate the properties of the proposed approach. It is first shown how these ODEs can be linearized to define the steps in the QL-ODE-P-spline estimation algorithm.
The ODE given in Eq. (2) can be linearized at iteration $i+1$ as follows:

$$\begin{aligned}\frac{d\widetilde{x}^{(i+1)}}{dt}(t) &= (\theta_1 - 2\theta_2 t)\left[\widetilde{x}^{(i+1)}(t)\right]^2 \\ &\approx 2\left(\theta_1 - 2\theta_2 t\right)\widetilde{x}^{(i)}(t)\,\widetilde{x}^{(i+1)}(t) - \left(\theta_1 - 2\theta_2 t\right)\left[\widetilde{x}^{(i)}(t)\right]^2.\end{aligned}$$

In the Van der Pol system, only the first differential equation has a nonlinear term. It can be linearized using:

$$\begin{aligned}\frac{d\widetilde{x}_1^{(i+1)}}{dt}(t) &= \theta\left(\widetilde{x}_1^{(i+1)}(t) - \frac{1}{3}\left[\widetilde{x}_1^{(i+1)}(t)\right]^3 - \widetilde{x}_2^{(i+1)}(t)\right) \\ &\approx \theta\left(1 - \left[\widetilde{x}_1^{(i)}(t)\right]^2\right)\widetilde{x}_1^{(i+1)}(t) - \theta\widetilde{x}_2^{(i+1)}(t) + \frac{2\theta}{3}\left[\widetilde{x}_1^{(i)}(t)\right]^3.\end{aligned}$$

These approximations are then used to build the penalty terms in Eqs. (5) and (6). The spline coefficients and the other parameters are jointly updated with the ODE-compliance parameters (using the EM-type approach by Schall (1991)) as in Algorithm 1.

4.1 Evaluation of estimation accuracy

An intensive simulation study is presented here to study the properties of the proposed methods. We have generated a large number of datasets using the ODE models in Eqs. (2) and (4) by perturbing the analytic or numerical solutions of the differential models by an additive zero mean Gaussian noise.
We aim to assess the quality of the estimates provided by the QL-ODE-P-spline framework. The performances of the proposed approach are compared with those achieved by a direct NLS analysis based on a Runge-Kutta numerical scheme.

4.1.1 First order ODE simulation results

In this section, the results of simulations based on Eq. (2) are discussed. The (analytic) state function has been computed with $\theta_1 = \theta_2 = 1$ and $x(0) = 1$. Gaussian noise with mean zero and variance $\sigma^2 = 0.045^2$ has been added to N values of the state function evaluated at time points uniformly sampled in $[0, 2]$. The data generation process has been repeated 500 times under four sample sizes $N = (20, 50, 100, 500)$.
Table 2 summarizes the estimation performances when $x(0)$ is unknown and provides the bias (in percentage), the averaged standard errors, the relative mean squared error and standard deviation of $\widehat{\theta}_1, \widehat{\theta}_2$ and $\widehat{\tau}$. In all the simulation settings we used cubic B-splines defined on 40 equidistant internal knots. The quality of the QL-ODE-P-spline estimates are compared with those of the Runge-Kutta based nonlinear least squares approach with initial state conditions approximated using smoothing splines (Wu et al., 2008).
The table clearly shows that the quasilinearized ODE-P-spline provides better estimates of the unknown ODE parameters than the NLS-based approach. The biases for all the parameter estimates are small and similar whatever the sample size. Similar results are obtained for the estimation of the standard deviations and the root mean squared errors of the parameter estimators.

[Table 2 about here.]

4.1.2 Van der Pol system simulation results

A simulation study was also performed with the Van der Pol system in Eq. (4). Five hundreds datasets of size $N = (50, 100, 500, 1000)$ have been obtained by adding a random Gaussian noise component with zero mean and variance $\tau^{-1} = \sigma^2 = 0.07^2$ to uniformly sampled values of the numerical solution to Eq. (4). The solution has been approximated using a Runge-Kutta scheme for values of t in $[0, 10]$ and considering initial value conditions $x_1(0) = x_2(0) = 1$ for $\theta = 1$. Table 3 compares the QL-ODE-P-splines method and the Runge-Kutta NLS approach in terms of estimation performances. The results of our smoothing-based procedures have been obtained using fourth order B-splines defined on 150 equidistant internal knots. The initial state conditions for the NLS-based approach have been approximated using smoothing splines.

Analogously to the previous simulation example, the QL-ODE-P-spline approach seems to ensure better performances and this becomes even more evident for moderate sample sizes. The biases and the root mean squared errors of the estimates of the ODE parameters and of the state function at $t = 0$ are quite small and tend to decrease with sample size.

[Table 3 about here.]

5 Two real(istic) examples

In this section the QL-ODE-P-spline procedure is illustrated by analyzing two real(isitc) data examples. As first application we focus on a simulated experiment inspired by the heartbeat phenomenological model proposed by dos Santos et al. (2004). In the second example, (a subset of) the well known Canadian lynx-hare abundance data (Elton and Nicholson, 1942; Holling, 1959) are analyzed by estimating the parameter and the solution of a Lotka-Volterra system of equations via QL-ODE-P-splines.

5.1 A realistic model for the analysis of heartbeat signals

The normal cardiac rhythm depends on the aggregate of cells in the right atrium defining the sino-atrial (SA) node which constitutes the normal pacemaker. The SA node generates electrical impulses that spread to the ventricules through the atrial musculature and conducting tissues, the AV node. The idea of treating the heart system using differential models dates back to the work by Van der Pol and Van der Mark (1928). In their pioneering paper the authors proposed to model the SA/AV system by a coupled set of electronic systems exhibiting relaxation oscillations. The Van der Pol and Van der Mark (1928) model represents a convenient description of the hearth dynamics due to its parametric simplicity and ability to cover complex periodicity. Successively, other systems have been proposed generalizing this original framework in order to better synthesize the biological complexity of the phenomenon under consideration (see Formaggia et al. (2010) for more details).
Among all the alternatives proposed in the specialized literature, we adopt the following system of coupled oscillators (dos Santos et al., 2004):

$$\begin{cases} \dfrac{dx_1}{dt}(t) = & x_2(t) \\ \dfrac{dx_2}{dt}(t) = & \kappa(x_1(t) - w_1)(x_1(t) - w_2)x_2(t) - b_1 x_1(t) + c_1(x_3(t) - x_1(t)) \\ \dfrac{dx_3}{dt}(t) = & x_4(t) \\ \dfrac{dx_4}{dt}(t) = & \kappa(x_3(t) - w_1)(x_3(t) - w_2)x_4(t) - b_2 x_3(t) + c_2(x_1(t) - x_3(t)) \end{cases}$$

where (x_1, x_2) and (x_3, x_4) describe the electrical activity in the SA (sino-atrial) and AV (atrio-ventricular) nodes respectively. Parameters b_1 and b_2 determine the normal mode frequencies and (w_1, w_2) characterize the non-linearity of the system. A central role is played by parameters (c_1, c_2) determining the kind of diffusivity described by the coupled oscillators. For $c_1 = c_2 = 0$ the SA node undergoes periodic regular oscillations, while for non-null $c_1 < c_2$ the system describes a bidirectional asymmetric coupling.
The right panels in Figure 3 depict the simulated measurements (dots) and the numerical solution of the system (dashed lines) together with the state function approximated via the QL-ODE-P-spline method (black full lines). The data have been obtained by adding a zero mean Gaussian noise to 200 uniformly sampled values of the first derivatives of the numerical solutions of the system defined by using ODE-parameters consistent with those in dos Santos et al. (2004). The system has been solved numerically for $x_1(0) = 1, x_2(0) = 1, x_3(0) = 1, x_4(0) = 1$ taking $t \in (0, 20)$. Table 4 shows the estimated parameters and their standard errors. In order to appropriately approximate the complexity of the observed

dynamics, fourth order B-splines built on 250 equidistant internal knots have been used in this example. From Figure 3 and Table 4 one can appreciate the quality of the performances achieved by the QL-ODE-P-spline procedure.

[Table 4 about here.]

[Figure 3 about here.]

5.2 Canadian lynx and snowshoe hare life cycles

In this section, we analyze the well known Canadian lynx and snowshoe hare data (Elton and Nicholson, 1942; Holling, 1959). The lynx and hare abundances have been collected from historical pelt-trading records of the Hudson Bay Company in a period between 1862 and 1930. Here, we consider a subset of the original sample analyzing the abundances observed between 1900 and 1920 (see figure 4). For these data the standard Lotka-Volterra system of ODEs can be estimated to obtain a meaningful description the life cycles of the two populations:

$$\begin{cases} \dfrac{dx_1}{dt}(t) = x_1(t)\left[\beta - \zeta x_2(t)\right] \\ \dfrac{dx_2}{dt}(t) = -x_2(t)\left[\delta - \eta x_1(t)\right] \end{cases}$$

where (x, y) indicate the prey (hare) and the predator (lynx) abundances respectively, β and η are the intrinsic growth rates of the prey and predator populations and (ζ, δ) are the predator-pray meeting and the predator death rates respectively.
The parameter estimates obtained with the quasilinearized ODE-P-splines are given in Table **??**. These results were computed using cubic B-splines defined on a set of 100 internal knots. The initial values for the ODE and precision parameters, $\boldsymbol{\theta}^{(0)} = (0.547, 0.028, 0.843, 0.026)$ and $\tau^{(0)} = 0.01$, were obtained using nonlinear least squares with x_1 and x_2 approximated using standard P-splines (Eilers and Marx, 1996) with smoothing parameter selected by generalized cross validation (Wahba, 1990). Figure 4 shows the approximated state functions and compares them with the raw measurements and the numerical solutions of the ODE system (black dashed lines). These numerical solutions were computed using a Runge-Kutta scheme for ODE parameters and initial values set at their QL-ODE-P-spline estimates.
Applying the proposed QL-ODE-P-spline method, a smoother consistent with the ODE model is estimated (Figure 4): the approximated state functions (solid gray lines) and the numerical solutions (black dashed lines) overlap. In our opinion, these estimates guarantee a good description of the data although the fitted state functions do not catch the trend observed in the years 1905 to 1910. This miss-fitting effect is a consequence of the poor flexibility of the Lotka-Volterra model. However we consider it a small price to be payed given the model parsimony and interpretability.
This fitting issue could also be related to the data collection process. Indeed, data before 1903 represent fur trading records while they were derived from questionnaires after 1903 (Zhang et al., 2007). This might have influenced the observations in the years between the two peaks. Finally, according to our experience, we advise to start the estimation process by taking a reasonably large $\gamma^{(0)}$ parameter if the research interest is focused on the estimation of the ODE parameters and the approximation of state functions consistent with the ODE system.

[Table 5 about here.]

[Figure 4 about here.]

6 Discussion

In this paper we have introduced the quasilinearized (QL-) ODE-P-spline approach as a tool for the estimation of the parameters defining (systems of) nonlinear ODEs and for the approximation of their state functions starting from a set of (noisy) measurements. This methodology is inspired by the generalized profiling framework of Ramsay et al. (2007). The estimation procedure exploits a penalized likelihood formulation to build a compromise between a B-spline based description of the observations and the B-spline collocation solution to the ODE. The estimation process alternates two steps: 1) estimation of the optimal spline coefficients given the other unknowns; 2) estimation of the ODE parameters treating the spline coefficients as nuisance parameters. The compromise between data smoothing and compliance to the differential model is tuned by an ODE-compliance parameter to be selected in an upper optimization level. Dealing with nonlinear ODEs, the application of the original proposal of Ramsay et al. (2007) is demanding as it implies a nonlinear least squares step and the use of the implicit function theorem both for point and interval estimates. This is a consequence of the non explicit relationship between the spline coefficient estimates and the other unknowns.

The introduction of a quasilinearization step overcomes these hitches leading to a valuable simplification of the estimation procedure. The quasilinearization of the ODE-based penalty term (Bellman and Kalaba, 1965) permits to analytically link the spline coefficients to the differential parameters. This makes possible to select the ODE-compliance parameters using standard approaches and simplifies the approximation of the var-cov of the unknown parameters. Indeed, within our quasilinearized spline based framework, the estimation process reduces to a conditionally linear problem for the optimization of the spline coefficients. We suggest to view the ODE-compliance parameter as the ratio of the noise and of the penalty variances and to estimate it using the EM-type approach proposed by Schall (1991). The quasilinearized estimation process requires initial values $\boldsymbol{\alpha}^{(0)}$ for the spline coefficients (Algorithm 1). Good initial spline coefficients can be obtained by smoothing the raw data with standard P-splines (Eilers and Marx, 1996) even if, according to our experience, the choice of $\boldsymbol{\alpha}^{(0)}$ hardly influences the quality of the final estimates (see Section 3).

The performances of the proposed method have been evaluated through intensive simulations (Section 4). Two simulation studies have been conducted by generating data from a simple first order ODE in the first case and by considering a relaxation oscillator governed by the Van der Pol equation in the second case. The proposed approach has been compared with the direct NLS estimation procedure based on a Runge-Kutta numerical scheme. For the latter, the required initial values for the state function were approximated using smoothing splines. The simulation results confirm that our approach ensures better performances than the ones provided by a Runge-Kutta NLS approach.

In Section 5, the quasilinearized spline based methodology has been applied in two real(istic) data analyzes. In the first example we have analyzed a set of 150 synthetic SA/AV voltage measurements generated perturbing with an additive Gaussian noise the first derivatives of the (numerical) solutions of an unforced system of coupled Van der Pol equations (dos Santos et al., 2004). In the second example we have analyzed the (yearly) lynx-hare abundances (MacLulich, 1937; Elton and Nicholson, 1942) recorded in Canada between 1900 and 1920. In this case, we used a QL-ODE-P-spline approach with a penalty based on a Lotka-Volterra system.

Our future research will focus on possible extensions of the QL-ODE-P-spline framework. First of all, we think that the linearized ODE penalty and the profiling estimation settings can be adapted to hierarchical settings with random ODE parameters. Second, the presented approach can be also adapted to analyze (noisy) realizations of dynamic systems evolving in space and time and described by nonlinear partial differential equations (PDEs). Finally, the introduction of a quasilinearization step should facilitate the generalization of the approach proposed by Jaeger and Lambert (2013) to analyze nonlinear differential systems in a Bayesian framework.

Acknowledgements The authors acknowledge financial support from IAP research network P7/06 of the Belgian Government (Belgian Science Policy), and from the contract 'Projet d'Actions de Recherche Concertées' (ARC) 11/16-039 of the 'Communauté française de Belgique', granted by the 'Académie universitaire Louvain'.

Conflict of Interest

The authors have declared no conflict of interest.

Appendix

A. Construction of the quasilinearized penalty matrix in a general system of nonlinear ODEs

In this appendix we define the matrix $\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$, the vector $\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ and the constant $\boldsymbol{l}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ involved in the linearized ODE-penalty. Consider a general system of ODEs:

$$\begin{cases} \dfrac{dx_1}{dt}(t) - f_1(t,\boldsymbol{x},\boldsymbol{\theta}) = 0 \\ \quad\vdots \\ \dfrac{dx_d}{dt}(t) - f_d(t,\boldsymbol{x},\boldsymbol{\theta}) = 0 \end{cases} \tag{A.1}$$

where $f_j(t,\boldsymbol{x},\boldsymbol{\theta})$ is a known function of the state variables and of the unknown ODE parameters $\boldsymbol{\theta}$ that we consider fixed over time. With quasilinearization, given the estimate $\widetilde{\boldsymbol{x}}^{(i)}$ at step i, Eq. (A.1) suggests to require that $\widetilde{\boldsymbol{x}}^{(i+1)}$ checks:

$$\begin{cases} \dfrac{d\widetilde{x}_1^{(i+1)}}{dt}(t) - \displaystyle\sum_{k=1}^{d}\widetilde{x}_k^{(i+1)}\frac{\partial f_1}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) - f_1\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) + \sum_{k=1}^{d}\widetilde{x}_k^{(i)}\frac{\partial f_1}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) = 0 \\ \quad\vdots \\ \dfrac{d\widetilde{x}_d^{(i+1)}}{dt}(t) - \displaystyle\sum_{k=1}^{d}\widetilde{x}_k^{(i+1)}\frac{\partial f_d}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) - f_d\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) + \sum_{k=1}^{d}\widetilde{x}_k^{(i)}\frac{\partial f_d}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) = 0 \end{cases}$$

If $\boldsymbol{\alpha}_j$ denotes the vector of spline coefficients to be estimated in the jth state function, the updated jth penalty term at iteration $i+1$ is given by:

$$\begin{aligned} PEN_j\left(\boldsymbol{\alpha}_j^{(i+1)}|\boldsymbol{\alpha}_j^{(i)}\right) &\approx \int_0^T \left(\frac{d\widetilde{x}_j^{(i+1)}}{dt}(t) - f_j\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\right. \\ &- \left.\sum_{k=1}^{d}\left(\widetilde{x}_k^{(i+1)} - \widetilde{x}_k^{(i)}\right)\frac{\partial f_j}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\right)^2 dt, \\ &\approx \int_0^T \left(\left[\boldsymbol{\alpha}_j^{(i+1)}\right]^\top \boldsymbol{b}_j^{(1)}(t)\right. \\ &- \left.\sum_{k=1}^{d}\left[\boldsymbol{\alpha}_k^{(i+1)}\right]^\top \boldsymbol{b}_j^{(0)}(t)\frac{\partial f_i}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) - v_j(t)\right)^2 dt \end{aligned}$$

with $v_j(t) = f_j\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right) - \displaystyle\sum_{k=1}^{d}\widetilde{x}_k^{(i)}(t)\frac{\partial f_j}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)$ and $\boldsymbol{b}_j^{(0)}(t)$, $\boldsymbol{b}_j^{(1)}(t)$ are, respectively, K_j-dimensional vectors of B-spline functions and their derivative such that $\widetilde{x}_j^{(i)}(t) = \left[\boldsymbol{b}_j^{(0)}(t)\right]^\top \boldsymbol{\alpha}_j^{(i)}$, $\dfrac{d\widetilde{x}_j^{(i)}}{dt}(t) =$

$\left[\boldsymbol{b}_j^{(1)}(t)\right]^\top \boldsymbol{\alpha}_j^{(i)}$. The jth element of the penalty can be computed as:

$$\begin{aligned}
PEN_j &= \left[\boldsymbol{\alpha}_j^{(i+1)}\right]^\top \left(\int \boldsymbol{b}_j^{(1)}(t)\left[\boldsymbol{b}_j^{(1)}(t)\right]^\top dt\right)\boldsymbol{\alpha}_j^{(i+1)} \\
&- \left[\boldsymbol{\alpha}_j^{(i+1)}\right]^\top \left(\int \frac{\partial f_j}{\partial x_1}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_j^{(1)}(t)\left[\boldsymbol{b}_1^{(0)}(t)\right]^\top dt\cdots,\right. \\
&\left. \int \frac{\partial f_j}{\partial x_d}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_j^{(1)}(t)\left[\boldsymbol{b}_d^{(0)}(t)\right]^\top dt\right)\boldsymbol{\alpha}^{(i+1)} \\
&- \left[\boldsymbol{\alpha}^{(i+1)}\right]^\top \begin{pmatrix} \int \frac{\partial f_j}{\partial x_1}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_1^{(0)}(t)\left[\boldsymbol{b}_j^{(1)}(t)\right]^\top dt \\ \vdots \\ \int \frac{\partial f_j}{\partial x_d}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_d^{(0)}(t)\left[\boldsymbol{b}_j^{(1)}(t)\right]^\top dt \end{pmatrix}\boldsymbol{\alpha}_j^{(i+1)} \\
&+ \left[\boldsymbol{\alpha}^{(i+1)}\right]^\top \begin{pmatrix} c_{j_{11}} & c_{j_{12}} & \cdots & c_{j_{1d}} \\ c_{j_{21}} & \cdots & \cdots & c_{j_{2d}} \\ \vdots & \vdots & \ddots & \vdots \\ c_{j_{d1}} & \cdots & \cdots & c_{j_{dd}} \end{pmatrix}\boldsymbol{\alpha}^{(i+1)} \\
&+ 2\left[\boldsymbol{\alpha}^{(i+1)}\right]^\top \begin{pmatrix} \int v_j(t)\frac{\partial f_j}{\partial x_1}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_j^{(0)}(t)dt \\ \vdots \\ \int v_j(t)\frac{\partial f_j}{\partial x_d}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_j^{(0)}(t)dt \end{pmatrix} \\
&- 2\left(\int v_j(t)\,\boldsymbol{b}_j^{(1)}(t)dt\right)^\top \boldsymbol{\alpha}^{(i+1)} + \int v_j^2(t)dt,
\end{aligned} \tag{A.2}$$

with

$$c_{j_{kl}} = \int \frac{\partial f_j}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\frac{\partial f_j}{\partial x_l}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_k^{(0)}(t)\left[\boldsymbol{b}_l^{(0)}(t)\right]^\top dt.$$

The overall penalty term $PEN^{(i+1)} = \sum_{j=1}^d \gamma_j PEN_j^{(i+1)}$, at step $i+1$ is therefore equal to:

$$PEN^{(i+1)} = \left[\boldsymbol{\alpha}^{(i+1)}\right]^\top \boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)\boldsymbol{\alpha}^{(i+1)} + 2\left[\boldsymbol{\alpha}^{(i+1)}\right]^\top \boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right) + l\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right).$$

The matrix $\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ has dimension $K \times K$, is symmetric and block-diagonal. The block (k,l) has dimension $K_k \times K_l$ and is computed as:

$$\begin{aligned}
&\left[\delta_{kl}\gamma_k \int \boldsymbol{b}_k^{(1)}(t)\left[\boldsymbol{b}_k^{(1)}(t)\right]^\top dt - \gamma_k \int \frac{\partial f_j}{\partial x_k}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_k^{(1)}(t)\left[\boldsymbol{b}_l^{(0)}(t)\right]^\top dt\right. \\
&\left. -\gamma_l \int \frac{\partial f_j}{\partial x_l}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_l^{(0)}(t)\left[\boldsymbol{b}_k^{(1)}(t)\right]^\top dt + \sum_{i=1}^d \gamma_i c_{i_{kl}}\right],
\end{aligned}$$

where $k,\ l \in \{1,\ldots,d\}$ and $\delta_{kl} = 1$ if $k = l$ and zero otherwise. The vector $\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ has length K with kth component given by

$$\sum_{j=1}^d \gamma_j\left(\int v_j(t)\frac{\partial f_j}{\partial x_1}\left(t,\widetilde{\boldsymbol{x}}^{(i)},\boldsymbol{\theta}\right)\boldsymbol{b}_k^{(0)}(t)\,dt\right) - \gamma_k \int v_k(t)\,\boldsymbol{b}_k^{(1)}(t)dt.$$

Finally the constant $l\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)$ in the penalty is equal to $\sum_{j=1}^{d}\gamma_j\int v_j^2\,(t)dt$.

B Hessian of $H(\boldsymbol{\theta}|\boldsymbol{y},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\tau}})$

In Section 3 we suggest to approximate the sampling variation of the unknown ODE parameters using the Fisher information matrix (FIM) at convergence of the QL-ODE-P-spline algorithm. This requires the computation of the Hessian matrix of:

$$H(\boldsymbol{\theta}|\boldsymbol{y},\widehat{\boldsymbol{\alpha}},\widehat{\boldsymbol{\tau}})=\sum_{j\in\mathcal{J}}\left\{-\frac{\widehat{\tau}_j}{2}\left\|\boldsymbol{y}_j-\widetilde{\boldsymbol{x}}_j\right\|^2\right\},$$

with $\widetilde{\boldsymbol{x}}_j=\boldsymbol{B}_j\widehat{\boldsymbol{\alpha}}_j(\widehat{\boldsymbol{\theta}},\widehat{\boldsymbol{\tau}},\widehat{\boldsymbol{\gamma}},\boldsymbol{y})$. As shown before, the quasilinearization of the penalty term makes possible to work out an analytic expression for the vector of spline coefficients and this greatly simplifies the computation of $\partial^2 H/\partial\boldsymbol{\theta}^2$.

To simplify our notation assume that one is only interested in the approximation of a single state function ($\mathcal{J}=1$). Suppose that at iteration $i+1$ the QL-ODE-P-spline algorithm reaches convergence. At convergence the optimal vector of spline coefficients can be written as (with the hat symbols omitted):

$$\boldsymbol{\alpha}^{(i+1)}=\left(\boldsymbol{G}+\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)\right)^{-1}\left(\boldsymbol{g}-\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)\right),$$

From our previous discussion it clearly appears that the penalty matrix $\boldsymbol{R}$ and the penalty vector $\boldsymbol{r}$ do not depend on the current spline coefficients but only on the previous ones. It is then possible to show that the $(k,l)-$element of the second derivative matrix of H can be computed as:

$$\frac{\partial^2 H}{\partial\theta_k\theta_l}=-\widehat{\tau}\left[-\boldsymbol{y}^\top\frac{\partial^2\widetilde{\boldsymbol{x}}}{\partial\theta_k\partial\theta_l}+\left(\frac{\partial^2\widetilde{\boldsymbol{x}}}{\partial\theta_k\partial\theta_l}\right)^\top\widetilde{\boldsymbol{x}}+\left(\frac{\partial\widetilde{\boldsymbol{x}}}{\partial\theta_k}\right)^\top\frac{\partial\widetilde{\boldsymbol{x}}}{\partial\theta_l}\right]. \tag{B.1}$$

Let $\boldsymbol{C}=\left(\boldsymbol{G}+\boldsymbol{R}(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)})\right)^{-1}$ and $\boldsymbol{g}=\widehat{\tau}\boldsymbol{B}^\top\boldsymbol{y}$, then the first and second derivatives of the approximated state function (at convergence) w.r.t. $\boldsymbol{\theta}$ can be computed as:

$$\frac{\partial\widetilde{\boldsymbol{x}}}{\partial\theta_k}=-\boldsymbol{B}\boldsymbol{C}\frac{d\boldsymbol{R}}{d\theta_k}\boldsymbol{C}\left(\widehat{\tau}\boldsymbol{B}^\top\boldsymbol{y}-\boldsymbol{r}\right)+\boldsymbol{B}\boldsymbol{C}\left(-\frac{d\boldsymbol{r}}{d\theta_k}\right), \tag{B.2}$$

$$\begin{aligned}\frac{\partial^2\widetilde{\boldsymbol{x}}}{\partial\theta_k\partial\theta_l}=&\boldsymbol{B}\boldsymbol{C}\frac{d\boldsymbol{R}}{d\theta_l}\boldsymbol{C}\frac{d\boldsymbol{R}}{d\theta_k}\boldsymbol{C}\left(\widehat{\tau}\boldsymbol{B}^\top\boldsymbol{y}-\boldsymbol{r}\right)-\boldsymbol{B}\boldsymbol{C}\frac{d^2\boldsymbol{R}}{d\theta_k d\theta_l}\boldsymbol{C}\left(\widehat{\tau}\boldsymbol{B}^\top\boldsymbol{y}-\boldsymbol{r}\right)\\&+\boldsymbol{B}\boldsymbol{C}\frac{d\boldsymbol{R}}{d\theta_k}\boldsymbol{C}\frac{d\boldsymbol{R}}{d\theta_l}\boldsymbol{C}\left(\widehat{\tau}\boldsymbol{B}^\top\boldsymbol{y}-\boldsymbol{r}\right)-\boldsymbol{B}\boldsymbol{C}\frac{d\boldsymbol{R}}{d\theta_k}\boldsymbol{C}\left(-\frac{d\boldsymbol{r}}{d\theta_l}\right)\\&-\boldsymbol{B}\boldsymbol{C}\frac{d\boldsymbol{R}}{d\theta_l}\boldsymbol{C}\left(-\frac{d\boldsymbol{r}}{d\theta_k}\right)+\boldsymbol{B}\boldsymbol{C}\left(-\frac{d^2\boldsymbol{r}}{d\theta_k d\theta_l}\right).\end{aligned} \tag{B.3}$$

The derivatives of the penalty matrix $\boldsymbol{R}$ and vector $\boldsymbol{r}$ can be easily computed using the fact that they only depend on the values of the spline coefficients at the previous step (see Appendix A).

Simulations suggest that the results obtained with the sophisticated pseudo δ-method by Ramsay et al. (2007) are very similar. Table 6 compares the (average) standard errors of the ODE-parameters computed from the FIM of the QL-ODE-P-splines at convergence with those computed via the pseudo δ-method in a generalized profiling framework. The comparison is based on the simulation studies discussed in Section 4 with compliance parameters for the generalized profiling approach set at their estimated values following the QL-ODE-P-spline procedure.

[Table 6 about here.]

C Including state conditions in the estimation process

A (system of) differential equation(s) can have a unique solution if condition(s) are imposed on the value(s) of the state function(s) and/or of one of its derivatives at a given time point(s). Initial and boundary value conditions are usually distinguished. In what follows, for the sake of brevity, we generically refer to them as "state conditions".
In many real data analyzes the state conditions are not known in advance and need to be estimated from the data. In some cases, these conditions can arise as characteristic of the observed dynamics or as theoretical constraints for the final fit. A real data example is discussed in Frasso et al. (2013) where the authors analyze a set of call option prices. In this case the non-arbitrage state conditions need to be satisfied by the approximated state function in order to obtain reasonable estimates. If available, such information can easily be included in the estimation process dealing with linear differential equations (Frasso et al., 2013). The same is true after quasilinearization of a nonlinear system.
Consider the general differential problem given by:

$$\begin{cases} \dfrac{d\boldsymbol{x}}{dt}(t) &= \boldsymbol{f}(t, \boldsymbol{x}, \boldsymbol{\theta}), \\ \boldsymbol{x}(t_0) &= \boldsymbol{s}_{t_0} \text{ for } t_0 \in \mathcal{T}_0, \end{cases}$$

where $\mathcal{T}_0$ denotes the subset of observed times where $\boldsymbol{x}$ is forced to be equal to $\boldsymbol{s}_{t_0}$. Using the B-spline approximation, these conditions become:

$$\boldsymbol{S}\boldsymbol{\alpha}^{(i+1)} = \boldsymbol{s}_{t_0}, \tag{C.1}$$

where $\boldsymbol{S}$ is a matrix with each row containing B-spline functions evaluated at a given $t_0 \in \mathcal{T}_0$. These state conditions can be included in the estimation process using two strategies.
The first one consists in treating them as an extra least squares penalty. Then, the optimal spline coefficient can be obtained by maximizing the following modified fitting criterion:

$$\begin{aligned} J\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\tau}, \boldsymbol{\alpha}^{(i)}, \boldsymbol{y}\right) &= \sum_{j \in \mathcal{J}} \left\{ \frac{n_j}{2} \log(\tau_j) \frac{\tau_j}{2} \left\| \boldsymbol{y}_j - \widetilde{x}_j^{(i+1)}(\boldsymbol{t}) \right\|^2 \right\} \\ &\quad - \frac{1}{2} PEN\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\alpha}^{(i)}, \boldsymbol{\gamma}, \boldsymbol{\theta}\right) \\ &\quad - \frac{\kappa}{2} \left(\boldsymbol{S}\boldsymbol{\alpha}^{(i+1)} - \boldsymbol{s}_{t_0}\right)^\top \left(\boldsymbol{S}\boldsymbol{\alpha}^{(i+1)} - \boldsymbol{s}_{t_0}\right), \end{aligned} \tag{C.2}$$

where κ is set at an arbitrary large value. Note that with such an extra penalty, the state condition is imposed only in a least square sense. With the fitting criterion in Eq. (C.2), the optimal spline coefficients at iteration $(i+1)$ become:

$$\begin{aligned} \boldsymbol{\alpha}^{(i+1)} &= \underset{\boldsymbol{\alpha}}{\text{argmax}}\, J\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\theta}, \boldsymbol{\tau}, \boldsymbol{\gamma}, \boldsymbol{\alpha}^{(i)}\boldsymbol{y}\right) \\ &= \left(\boldsymbol{G} + \boldsymbol{R}\left(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}^{(i)}\right)\kappa \boldsymbol{S}^\top \boldsymbol{S}\right)^{-1} \left(\boldsymbol{g} - \boldsymbol{r}\left(\boldsymbol{\theta}, \boldsymbol{\gamma}, \boldsymbol{\alpha}^{(i)}\right) + \kappa \boldsymbol{S}^\top \boldsymbol{s}_{t_0}\right). \end{aligned}$$

If κ is equal to zero, then one gets back the estimation procedure discussed before, while for κ tending to infinity, the state functions are forced to meet the imposed conditions.
The second approach consists in forcing the smooth estimates of the state function to check the prescribed conditions by the introduction of Lagrange multipliers. Starting from Eq. (7) the Lagrange function for the

constrained optimization problem is:

$$\begin{aligned}\mathcal{L}\left(\boldsymbol{\alpha}^{(i+1)},\boldsymbol{\omega}|\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\tau},\boldsymbol{\alpha}^{(i)},\boldsymbol{y}\right) &= \sum_{j\in\mathcal{J}}\left\{\frac{n_j}{2}\log\left(\tau_j\right)-\frac{\tau_j}{2}\left\|\boldsymbol{y}_j-\widetilde{x}_j^{(i+1)}\left(\boldsymbol{t}\right)\right\|^2\right\}\\ &- \frac{1}{2}PEN\left(\boldsymbol{\alpha}^{(i+1)}|\boldsymbol{\alpha}^{(i)},\boldsymbol{\gamma},\boldsymbol{\theta}\right)\\ &- \boldsymbol{\omega}^\top\left(\boldsymbol{S}\boldsymbol{\alpha}^{(i+1)}-\boldsymbol{s}_{t_0}\right), \qquad \text{(C.3)}\end{aligned}$$

where $\boldsymbol{\omega}$ is the vector of Lagrange multipliers. Following Currie (2013), the maximization of $\mathcal{L}$ can be obtained by solving:

$$\begin{pmatrix} \boldsymbol{G}+\boldsymbol{R}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right) & \boldsymbol{S}^\top \\ \boldsymbol{S} & \boldsymbol{0}\end{pmatrix}\begin{pmatrix}\boldsymbol{\alpha}^{(i+1)}\\ \boldsymbol{\omega}\end{pmatrix}=\begin{pmatrix}\boldsymbol{g}-\boldsymbol{r}\left(\boldsymbol{\theta},\boldsymbol{\gamma},\boldsymbol{\alpha}^{(i)}\right)\\ \boldsymbol{s}_{t_0}\end{pmatrix}.$$

Whatever the selected approach to force the state conditions, the estimates of the ODE parameters $\boldsymbol{\theta}$ are updated using Eq. (8).

References

Bellman, R. and Kalaba, R. (1965). *Quasilinearization and nonlinear boundary-value problems*. Modern analytic and computational methods in science and mathematics. American Elsevier Pub. Co.

Biegler, L., Damiano, J., and Blau, G. (1986). Nonlinear parameter estimation: A case study comparison. *AIChE Journal*, 32(1):29–45.

Currie, I. D. (2013). Smoothing constrained generalized linear models with an application to the Lee-Carter model. *Statistical Modelling*, 13(1):69–93.

Dierckx, P. (1995). *Curve and Surface Fitting with Splines*. Oxford University Press.

dos Santos, A. M., Lopes, S. R., and Viana, R. L. (2004). Rhythm synchronization and chaotic modulation of coupled van der pol oscillators in a model for the heartbeat. *Physica A: Statistical Mechanics and its Applications*, 338(34):335 – 355.

Eilers, P. and Marx, B. (1996). Flexible smoothing with B-splines and penalties. *Statistical Science*, 11:89–121.

Eilers, P. and Marx, B. (2010). Splines, knots, and penalties. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(6):637–653.

Elton, C. and Nicholson, M. (1942). The ten-year cycle in numbers of the lynx in canada. *Journal of Animal Ecology*, 11(2):pp. 215–244.

Formaggia, L., Quarteroni, A., and Veneziani, A. (2010). *Cardiovascular Mathematics: Modeling and simulation of the circulatory system*. MS&A. Springer.

Frasso, G., Jaeger, J., and Lambert, P. (2013). Estimation and approximation in multidimensional dynamics. Technical report, Université de Liège.

Golub, G. H. and Ortega, J. M. (1992). *Scientific Computing and Differential Equations*. Academic Press, New York and London.

Hastie, T. J. and Tibshirani, R. J. (1990). *Generalized additive models*. London: Chapman & Hall.

Holling, C. S. (1959). The components of predation as revealed by a study of small mammal predation predation of the european pine sawfly. *Canadian Entomologist*, 91:293–320.

Hotelling, H. (1927). Differential equations subject to error, and population estimates. *Journal of the American Statistical Association*, 22(159):283–314.

Jaeger, J. and Lambert, P. (2013). Bayesian P-spline estimation in hierarchical models specified by systems of affine differential equations. *Statistical Modelling*, 13(1):3–40.

MacLulich, D. A. (1937). *Fluctuations in the numbers of varying hare (Lepus americanus)*. University of Toronto studies: Biological series. The University of Toronto press.

Ramsay, J. O., Hooker, G., Campbell, D., and Cao, J. (2007). Parameter estimation for differential equations: a generalized smoothing approach. *Journal of the Royal Statistical Society, Series B*, 69:741–796.

Rodriguez-Fernandez, M., Egea, J., and Banga, J. (2006). Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. *BMC Bioinformatics*, 7(1):483.

Schall, R. (1991). Estimation in generalized linear models with random effects. *Biometrika*, 78(4):719–727.

Van der Pol, B. (1926). On relaxation-oscillations. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 7(2):978–992.

Van der Pol, B. and Van der Mark, J. (1928). The heartbeat considered as a relaxation oscillation, and an electrical model of the heart. *Philosophical Magazine Series 7*, 6(38):763–775.

Wahba, G. (1990). *Spline models for observational data*, volume 59 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Wu, H., Zhu, H., Miao, H., and Perelson, A. S. (2008). Parameter identifiability and estimation of hiv/aids dynamic models. *Bulletin of Mathematical Biology*, 70(3):785–799.

Xue, H., Miaou, H., and Wu, H. (2010). Sieve estimation of constant and time-varying coefficients in nonlinear ordinary differential equation models by considering both numerical error and measurement error. *Ann. Statist.*, 38(4):2351–2387.

Zhang, Z., Tao, Y., and Li, A. Z. (2007). Factors affecting hare-lynx dynamics in the classic time series of the Hudson Bay Company, Canada. *Climate Research*, 34:83–89.

List of Figures

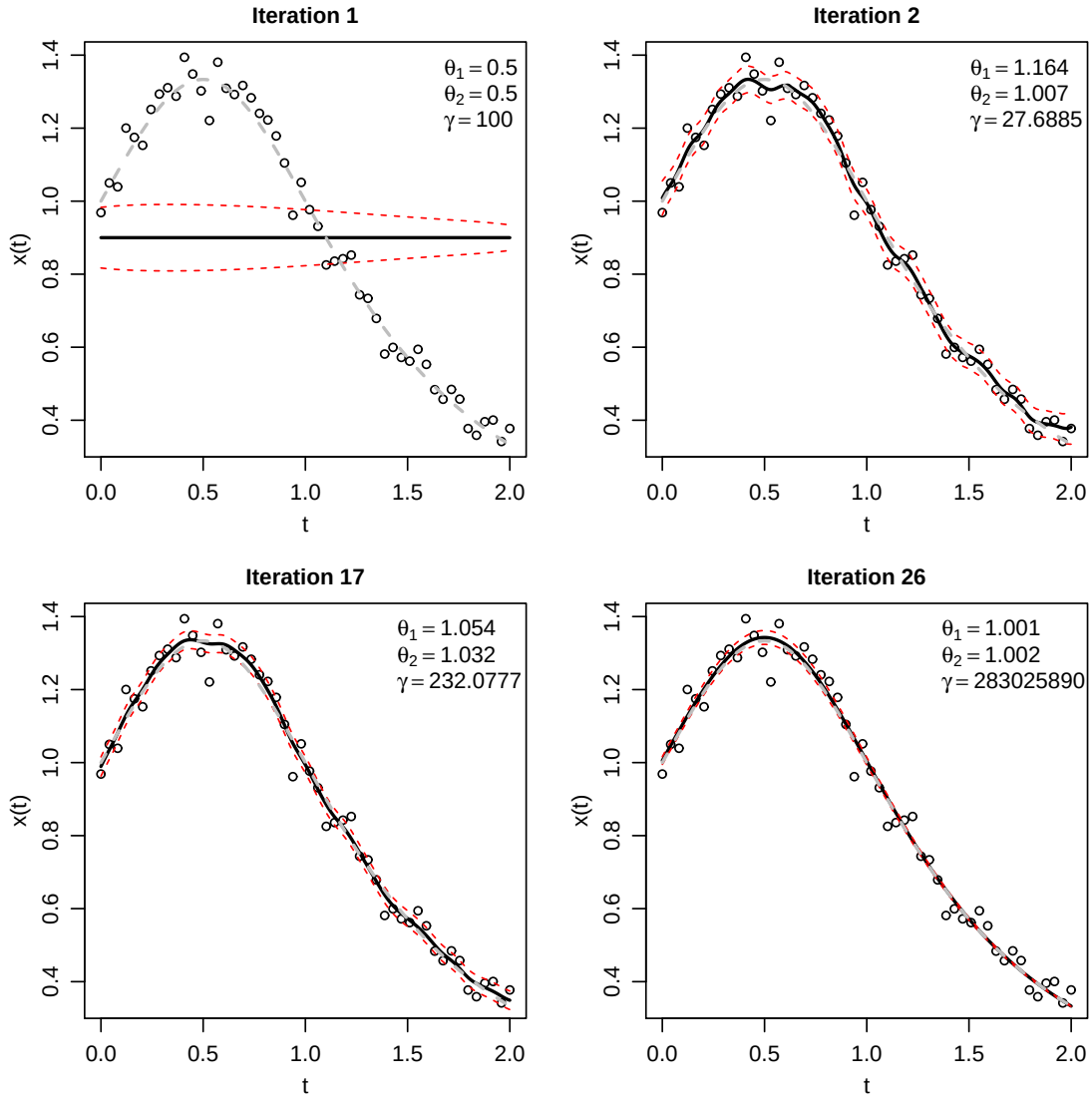

Figure 1 Selected sequence of estimation and approximation steps for the first order ODE model in Eq. (2). Simulated observations are represented by dots while the solid curves correspond to the approximated state functions, the dashed gray lines represent the exact solution to the differential problem and the dashed red ones indicate the approximated (± 2) standard error bands for the extracted state function. The ODE parameters used to simulate the data were $\theta_1^* = 1;\ \theta_2^* = 1$. The current estimates of $\boldsymbol{\theta}$ and γ are indicated in the legends.

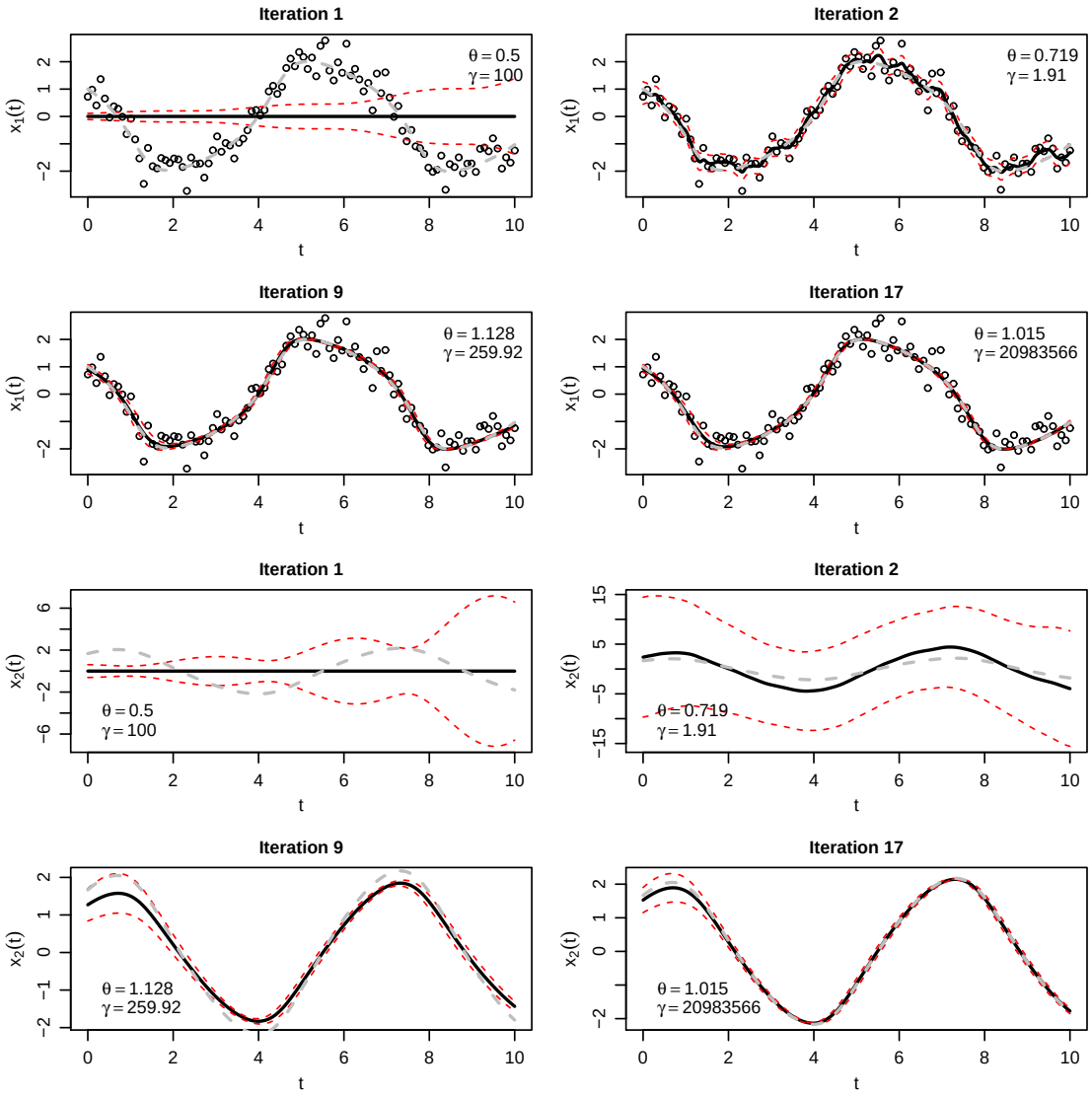

Figure 2 Selected sequence of estimation and approximation steps for the Van der Pol system in Eq. (4). Simulated observations for $x_1(t)$ ($x_2(t)$ is supposed not observed) are displayed by dots while the solid curves correspond to $\widetilde{\boldsymbol{x}}(t)$, the dashed gray ones represent the numerical solution to the differential problem (obtained using a Runge-Kutta scheme) and the dashed red ones indicate the approximated (± 2) standard error bands for the extracted state function. The ODE parameter used to simulate the data was $\theta^* = 1$. The current estimates of $\boldsymbol{\theta}$ and γ are indicated in the legends.

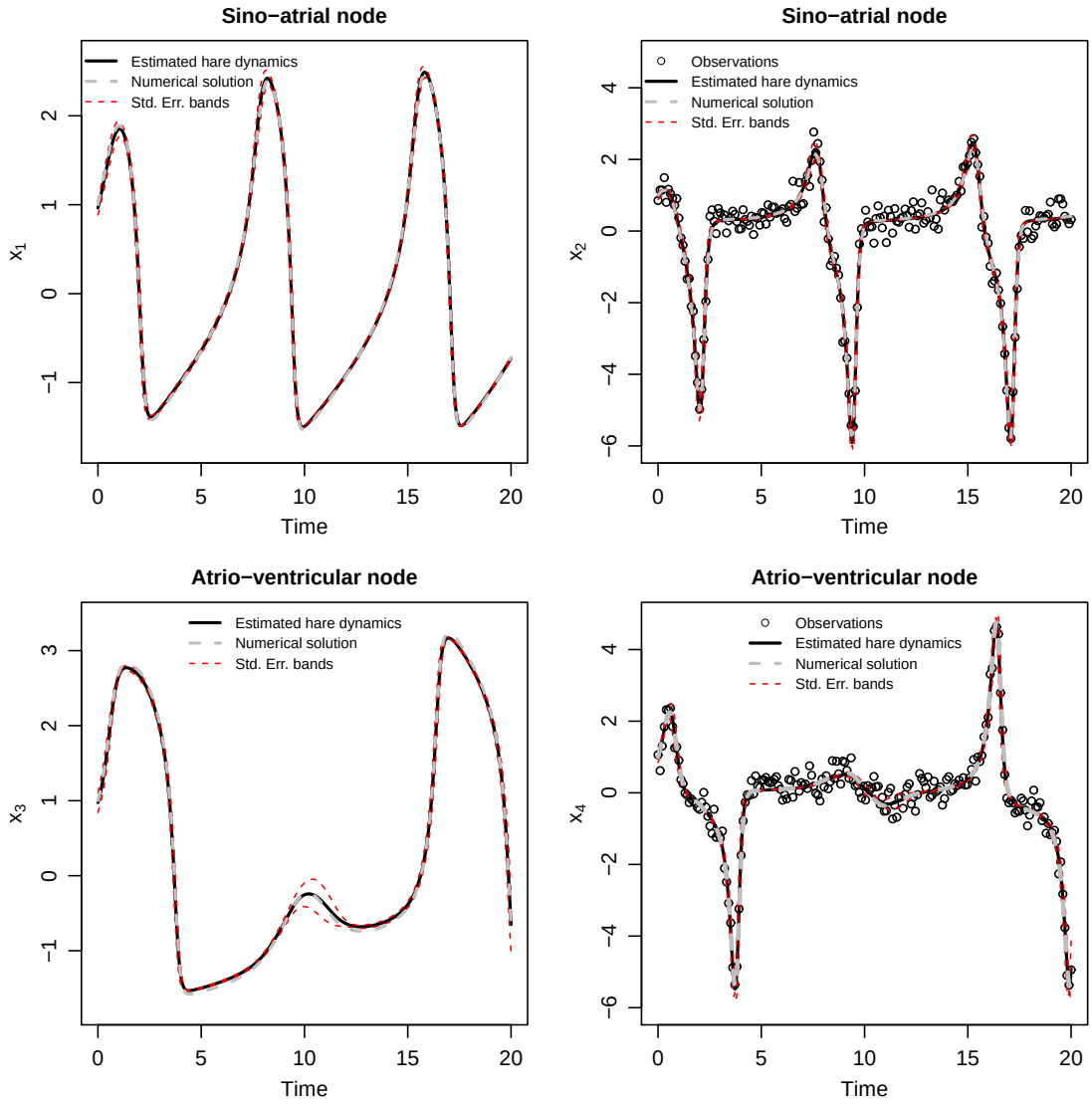

Figure 3 The first row of the plot matrix depicts a simulated SA node. Note that only the first derivative of the signal is observed (x_2). The second row shows the hypothetical voltage measured in the AV node. Again only the first derivative of the signal is observed (x_4). Two hundred measurements per node have been simulated.

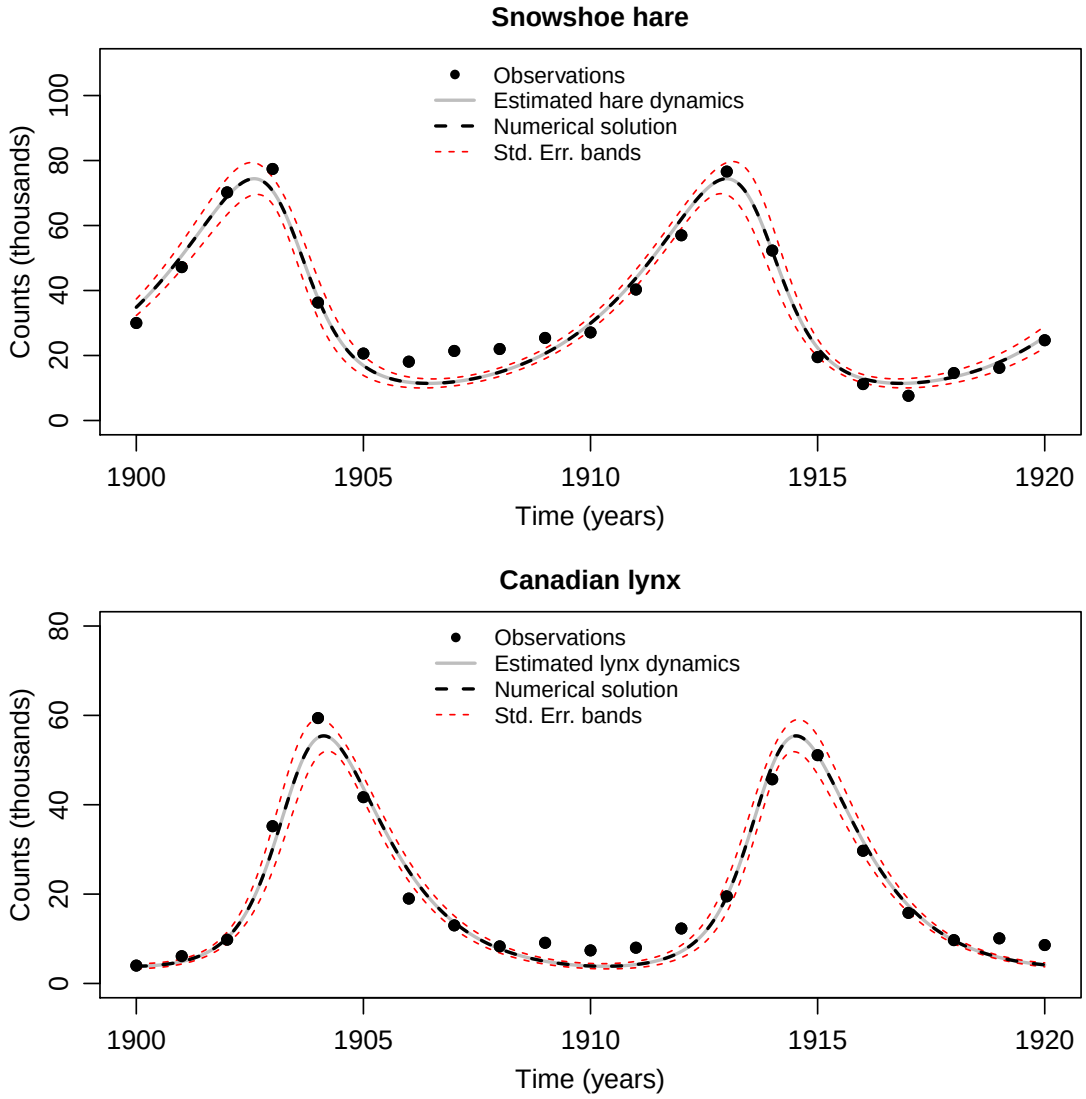

Figure 4 Raw data and estimates obtained for the Canadian lynx vs snowshoe hare predator-pray dynamics. In this example we take into account the yearly observations recorded between 1900 and 1920.

List of Tables

Table 1 Robustness of the QL-ODE-P-spline estimation approach to the choice of initial spline coefficients and ODE parameters. The first two rows of each sub-table show the (average) RMSEs between the approximated state function and the simulated measurements taking either constant initial values for the spline coefficients (Strategy 1) or a standard P-spline smoother (Strategy 2). Rows 3 and 4 report the (average) RMSEs computed with respect to the analytic/numerical solution of the ODE model. The remaining part of each sub-table reports the average number of iterations ($\overline{\#\text{ Iter}}$) required to achieve convergence.

	First order ODE				Van der Pol system			
	Initial $\boldsymbol{\alpha}$	$\boldsymbol{\theta}^*$	$0.5\boldsymbol{\theta}^*$	$2\boldsymbol{\theta}^*$	Initial $\boldsymbol{\alpha}$	$\boldsymbol{\theta}^*$	$0.5\boldsymbol{\theta}^*$	$2\boldsymbol{\theta}^*$
$\overline{\text{RMSE}(\boldsymbol{y})}$	Strategy 1	0.041	0.040	0.041	Strategy 1	0.042	0.040	0.041
	Strategy 2	0.042	0.040	0.041	Strategy 2	0.042	0.040	0.041
$\overline{\text{RMSE}(\boldsymbol{x})}$	Strategy 1	0.012	0.013	0.013	Strategy 1	0.011	0.013	0.013
	Strategy 2	0.011	0.013	0.013	Strategy 2	0.011	0.013	0.013
$\overline{\#\text{ Iter}}$	Strategy 1	16	22	21	Strategy 1	16	14	14
	Strategy 2	18	23	21	Strategy 2	16	14	14

Table 2 Simulation results for the first order ODE example. For the two approaches, the bias, the root mean squared errors, the standard deviations and the average standard errors of the estimated parameters (ODE and initial values) are provided. In the last row of the right part of the table, the (average) optimal ODE-compliance parameters are shown.

		Runge-Kutta NLS				**QL-ODE-P-splines**			
	Parameter	$N = 20$	$N = 50$	$N = 100$	$N = 500$	$N = 20$	$N = 50$	$N = 100$	$N = 500$
Bias	θ_1	1.018%	0.165%	0.046%	0.093%	-0.244%	0.038%	0.003%	-0.034%
	θ_2	0.999%	0.173%	0.156%	0.106%	-0.183%	0.021%	0.024%	-0.028%
	$x(0)$	-0.049%	-0.005%	0.039%	0.012%	0.082%	0.016%	0.023%	0.015%
	τ	33.265%	12.647%	4.927%	1.237%	32.198%	9.294%	4.724%	0.892%
RMSE	θ_1	1.34E-01	9.14E-02	6.74E-02	3.27E-02	7.85E-02	5.54E-02	3.60E-02	1.86E-02
	θ_2	9.77E-02	6.23E-02	4.53E-02	2.11E-02	5.65E-02	3.89E-02	2.53E-02	1.33E-02
	$x(0)$	4.40E-02	3.13E-02	2.31E-02	1.15E-02	2.48E-02	1.76E-02	1.19E-02	5.68E-03
	τ	6.04E-01	2.67E-01	1.55E-01	6.28E-02	1.27E-03	4.93E-04	3.22E-04	1.34E-04
Std dev	θ_1	1.33E-01	9.15E-02	6.75E-02	3.27E-02	7.86E-02	5.55E-02	3.60E-02	1.86E-02
	θ_2	9.72E-02	6.23E-02	4.53E-02	2.11E-02	5.65E-02	3.90E-02	2.54E-02	1.33E-02
	$x(0)$	4.40E-02	3.13E-02	2.32E-02	1.16E-02	2.48E-02	1.76E-02	1.19E-02	5.68E-03
	τ	5.05E-01	2.36E-01	1.47E-01	6.16E-02	5.47E-01	2.29E-01	1.54E-01	6.66E-02
	$\mathrm{E}(\sigma_{\theta_1})$	1.25E-01	9.24E-02	6.96E-02	3.26E-02	7.52E-02	5.23E-02	3.80E-02	1.77E-02
	$\mathrm{E}(\sigma_{\theta_2})$	9.00E-02	6.32E-02	4.67E-02	2.14E-02	5.38E-02	3.70E-02	2.68E-02	1.25E-02
	$\mathrm{E}(\gamma)$					2.52E+07	2.44E+07	2.41E+07	2.40E+07

Table 3 Simulation results with a Van der Pol system: the biases, root mean squared errors, standard deviations and (average) standard errors for parameter estimates using QL-ODE-P-splines and Runge-Kutta based nonlinear least squares procedures. In the last row of the right part of the table, the (average) optimal ODE-compliance parameters are shown.

		Runge-Kutta NLS				**QL-ODE-P-splines**			
	Parameter	$N = 50$	$N = 100$	$N = 500$	$N = 1000$	$N = 50$	$N = 100$	$N = 500$	$N = 1000$
Bias	θ	-0.378%	-0.210%	0.065%	0.012%	0.165%	-0.015%	-0.002%	-0.009%
	$x_1(0)$	0.146%	0.174%	0.011%	0.018%	-0.178%	-0.045%	0.015%	-0.021%
	$x_2(0)$	-0.009%	0.062%	-0.071%	0.049%	-0.166%	0.059%	0.034%	-0.002%
	τ	9.971%	5.424%	4.743%	-0.892%	7.495%	3.936%	0.699%	0.292%
RMSE	θ	6.73E-02	4.76E-02	7.48E-03	6.63E-03	3.28E-02	2.35E-02	1.01E-02	8.23E-03
	$x_1(0)$	3.83E-02	2.71E-02	1.02E-02	7.02E-03	3.07E-02	2.14E-02	9.23E-03	6.90E-03
	$x_2(0)$	2.45E-02	1.76E-02	7.08E-03	1.13E-02	5.83E-02	4.12E-02	1.82E-02	1.36E-02
	τ	2.58E-01	1.67E-01	6.18E-02	4.60E-02	2.42E-01	1.57E-01	6.67E-02	4.65E-02
Std dev	θ	6.73E-02	4.76E-02	7.46E-03	6.63E-03	3.28E-02	2.35E-02	1.01E-02	8.24E-03
	$x_1(0)$	3.83E-02	2.71E-02	1.02E-02	7.02E-03	3.07E-02	2.14E-02	9.24E-03	6.90E-03
	$x_2(0)$	2.45E-02	1.76E-02	7.05E-03	1.13E-02	5.84E-02	4.13E-02	1.82E-02	1.37E-02
	τ	2.39E-01	1.59E-01	6.13E-02	4.59E-02	2.30E-01	1.52E-01	6.64E-02	4.64E-02
	E(σ_θ)	2.38E-02	1.69E-02	7.47E-03	6.74E-03	2.90E-02	2.11E-02	9.67E-03	7.05E-03
	E(γ)					8.65E+07	8.74E+07	8.69E+07	8.58E+07

Table 4 ODE and precision parameter estimated for the observed coupled Van der Pol system (dos Santos et al., 2004). In the first column the true values of the ODE and precision parameters are indicated.

True parameter	**Estimates**	**S. E.**
$b_1 = 1.5$	1.514	0.018
$b_2 = 0.1$	0.120	0.022
$\kappa = -1.8$	-1.887	0.049
$w_1 = -0.2$	-0.197	0.012
$w_2 = 2$	1.972	0.020
$c_1 = 0.05$	0.044	0.016
$c_2 = 0.55$	0.537	0.023
$\tau = 30$	29.79	
$\widehat{\gamma}$	9.12E+03	

Table 5 Estimated parameters of the Lotka-Volterra system defined for the lynx-hare data example. The optimal ODE compliance parameters have been found equal to 2.71E+06.

Parameter	**Estimates**	**S. E.**
β	0.481	3.70E-02
ζ	0.025	2.00E-03
δ	0.927	7.60E-02
η	0.028	2.00E-03
$\widehat{\tau}$	3.863	

Table 6 Average standard errors of the ODE-parameters computed using the FIM of the QL-ODE-P-splines and the pseudo δ-method in a generalized profiling framework.

First Oredr ODE example

	QL-ODE-P-Splines				**Generalized Profiling**			
	N = 20	N = 50	N = 100	N=500	N = 20	N = 50	N = 100	N=500
θ_1	7.52E-02	5.23E-02	3.80E-02	1.77E-02	7.25E-02	5.12E-02	3.75E-02	1.72E-02
θ_2	5.38E-02	3.70E-02	2.68E-02	1.25E-02	6.78E-02	5.67E-02	2.60E-02	1.21E-02

Van der Pol example

	QL-ODE-P-Splines				**Generalized Profiling**			
	N = 50	N = 100	N = 500	N = 1000	N = 50	N = 100	N = 500	N = 1000
θ_1	2.90E-02	2.11E-02	9.67E-03	7.05E-03	2.76E-02	2.11E-02	1.01E-02	7.97E-03

