Ecole Polytechnique Federale de Lausanne

Applied Mathematics

Semester Project

ODE parameter estimation through a runner's model application

Author:
Lucien Roquette

Supervisor:
Professor A. C. Davison

25th July 2016

Acknowledgments

I would like to express my sincere gratitude to Prof. A. C. Davison for accepting to oversee me all along this project. I really appreciate that he shared his precious time with me by carefully supervising my writing, giving me useful comments and offering me his moral support for this research. I hope this report will be up to the great honour I had to work under his supervision. In addition, I would like to warmly thank the LMAM for providing me with the data, and particularly Mathieu Falbriad for sharing his experience of sports' measurements with me, being always available and even accepting to run himself for science!

Contents

Abstract

The monitoring of sports' performance has become increasingly popular in the past years. For instance, high precision measurement of runners' speed, foot pace and heart beat are available among others. Usually, coaches are interested in summary statistics of these data to assess the performance of their runners. However, it is uncommon to find a model that exploits the whole time series. The present report utilizes an ordinary differential equation (ODE) model that captures the runner's behavior all along his race. This model is based on physical principles, and depends on parameters related to the runner's physiological characteristics. Based on speed measurements along 200-meter races, this report explores two inference methods specific to the ODE model: the generalized profiling (GP) and the two-step method. Their efficiency will be compared to the standard non-linear least squares (NLS) method. Speed profiles have been sampled at high frequency, and present a correlation structure accounted for with an ARMA process. While the motivation to introduce the GP and the two-step methods are not strictly met for this runner's model, they present similar estimation quality as the NLS method and emphasize interesting features in terms of model fitting and mis-specification.

1 Introduction & Motivation

1.1 Motivation

The aim of this project is to estimate physiological parameters defining a differential model for a runner. The model is composed of a system of ODEs for the velocity $v(t)$ and the anaerobic energy $e(t)$. The system of ODEs used in this study was introduced in Aftaltion and Bonnans (2014), going further than the energetic modelling proposed in Keller (1974). This paper uses the ODE model in an optimal control framework to answer the following question: given a runner's physiological parameters, and a given distance, how can he minimize his running time? In Aftalion et al. (2016), these optimal velocity profiles were used to fit real data on 1500 m races in order to identify the corresponding physiological parameters. However, this procedure suffers from a lack of flexibility because the fitted model had been chosen only among optimal strategies. The more natural approach would be to fit the ODE model to the data, identify the physiological parameters involved and then generate an optimal velocity profile. From this approach emerged the motivation of studying parameter estimation from the ODE model. Moreover, an appropriate uncertainty quantification on physiological parameters and the fitted velocity profile could result in interesting outcomes. In particular, it could be relevant to generate optimal races using the estimated parameter's confidence ranges, and compare them to the confidence band on the velocity profile. This procedure could lead to an evaluation tool for runners' performance throughout their races.

1.2 Model description

The runner is modeled as a material point associated to an energy tank. This material point represents the runner's center of mass. The runner's state is defined according to: its velocity $v(t)$ and its anaerobic energy $e(t)$. For any time-dependent function $h(t)$, the notation $\dot{h}(t)$ is used for the derivative with respect to time. During the race, states'

evolution are ruled by the following ODE system:

$$\begin{cases} \dot{v}(t) = -v(t)/\tau + f(t), & v(0) = 0, \\ \dot{e}(t) = \sigma(e(t)) - f(t)v(t), & e(0) = e_0, \end{cases} \tag{1}$$

where $f(t)$ represents the propulsive force by unit of mass, $\sigma\{e(t)\}$ the energetic equivalent of the rate of oxygen uptake by unit of mass, and $t \in [0, T]$, with T the final time of the run. The velocity equation comes from the fundamental principle of dynamics. The energetic consumption equation is an energy conservation principle. A physical interpretation of each terms is:

- velocity equation: v/τ is a body friction force by unit of mass that limits the amount of propulsive force $f(t)$ inducing an increase of speed;
- energetic equation: anaerobic energy is consumed if the power spent -$f(t)v(t)$ is not offset by the creation of power from oxygen uptake $\sigma(e)$.

Anaerobic energy $e(t)$ plays the role of an energy tank for the system, only if we add the constraint $e(t) \geq 0$ for $t \in [0, T]$. The main aim of this model is to capture energy consumption along the race. From this energetic consumption, the model may predict interesting strategies and optimal races. These predictions are better suited for long distances, where the material point approximation is less important. For short distances, for instance up to 200 m, the stride rate, or length are much more important than strategy. In this context, the consideration of energetic savings does not lead to fruitful conclusions.

The energetic equivalent of oxygen uptake $\sigma\{e(t)\}$ is the amount of oxygen that is transformed in anaerobic energy along the race. The modelling of σ as a function of $e(t)$ was introduced in Aftaltion and Bonnans (2014). It was built to mimic experimental findings of Hanon and Thomas (2011). In the latter paper the evolution of the rate of oxygen uptake (denoted $\dot{V}O_2$) was recorded for 400, 800 and 1500 m and specifically trained runners. The study showed that there is a drop in $\dot{V}O_2$ for the last parts of the races, strongly correlated with tidal volume of lungs and maximal blood lactate concentration. The increase of $\dot{V}O_2$ at the beginning of the races was similar among all the runners, and the peak values for 400 m and 800/1500 m were respectively 94% and 100% of the known maximal rate of oxygen uptake, denoted $\dot{V}O_2^{\max}$. In Hanon and Thomas (2011), $\dot{V}O_2^{\max}$ is measured in $mL.kg^{-1}.min^{-1}$ while in Aftaltion and Bonnans (2014) and in the present report the energetic equivalent of this quantity will be used. Using the respiratory exchange rate $1L \longleftrightarrow 20kJ$, the energetic equivalent $\sigma(e)$ will be expressed in $J.kg^{-1}.s^{-1} = m^2.s^{-1}$. Thus the conversion factor is $60s/20kJ = 3$, i.e. the energetic equivalent is 3 times smaller than the measured $\dot{V}O_2^{\max}$. A typical value for the maximal value for $\sigma(e)$ is 21 $m^2.s^{-3}$.

The exact definition of the function $\sigma(e)$ is:

$$\sigma(e) = \begin{cases} \bar{\sigma}\frac{e}{e_0\gamma_2} + \sigma_f(1 - \frac{e}{e_0\gamma_2}), & \frac{e}{e_0} < \gamma_2, \\ \bar{\sigma}, & \frac{e}{e_0} \leq \gamma_1, \frac{e}{e_0} \geq \gamma_2, \\ \sigma_r + (\bar{\sigma} - \sigma_r)(1 - \frac{e}{e_0\gamma_1}), & \frac{e}{e_0} > \gamma_1, \end{cases} \tag{2}$$

where $\bar{\sigma}$ is the energetic equivalent of $\dot{V}O_2^{\max}$, σ_r and σ_f are respectively the rest and final values for oxygen uptake. The constants γ_1 and γ_2 define the critical value at which the regime of oxygen uptake changes. To be able to observe the three regimes, we must have $\gamma_2 < \gamma_1$, since the function $e(t)$ should be decreasing from $e(0) = e_0$. One should

think of $\sigma(t) = \sigma\{e(t)\}$ as a continuous piecewise linear function with three pieces that correspond to an increasing phase of oxygen uptake from σ_r, followed by a constant phase $\bar{\sigma}$, and a drop to σ_f. For a sprint effort the anaerobic energy is the main fuel, indeed $\sigma(e)$ stays in the first phase, i.e., much lower than $\bar{\sigma}$. Given the results in Hanon and Thomas (2011), it seems that parameters γ_1, γ_2 are likely to be similar among runners.

Thus, the most important physiological parameters involved in the dynamics of this model are $\bar{\sigma}$, e_0 and τ. In addition to these, the propulsive force $f(t)$ is limited by physiological constraints that could be considered as physiological parameters as well. For instance, a maximal propulsive force $f_{\max}$ that bounds $f(t)$, and a maximal instantaneous variation C of this force that bounds $|\dot{f}(t)|$, certainly exist.

Given the physiological parameters $\bar{\sigma}$, e_0, τ, C, and $f_{\max}$ it is possible to simulate an optimal race, and therefore find the control function $f(t)$ that generates the smallest final time T achievable given the physiological limitations of a runner. The details of this optimal control problem can be found in Aftaltion and Bonnans (2014) and Aftalion et al. (2016). In this report, the inverse problem of estimating runners' physiological parameters and propulsive force from the ODE model will be explored.

2 Description of the LMAM's data and methodology

2.1 Exploratory analysis

The Laboratory of Movement Analysis and Measurement (LMAM) provides us with speed recordings based on GPS technology. This equipment works at a sampling frequency of 10 Hz, and as a consequence there are a lot of data points even for a short distance. In Figure 1 three speed recording are displayed for 200 m runs, with roughly 400 points of observation for each subject. Oxygen uptake, $\dot{V}O_2$, measurements during an outdoor run (i.e., not on a treadmill) is difficult for practical reasons: the equipment must be carried by the runner. The measurement of anaerobic energy in the muscles is very invasive, and can only be done through laboratory experiments. The force applied on the floor $F(t)$ and the acceleration $\dot{v}(t)$ are the only additional measurements available. The runner's sole could be equipped with a sensor to record $F(t)$. However, the LMAM does not possess simultaneous measurements of $v(t)$ and $F(t)$ for outdoor races.

The LMAM also provides us with some indoor measurements of speed (deduced from the treadmill's speed), and of the force applied by the feet on the treadmill. These indoor experiments correspond to constant speed trials (see Figure 2). The runner's speed is a quantity that makes perfect sense in terms of the ODE model and of real measurements. It is relevant to better understand the link between the propulsive force $f(t)$ in the ODE model and the force $F(t)$ applied on the floor. Indeed, a clear relation between $f(t)$ and $F(t)$ would bring additional information for the fitting process. In Figure 2 (Right), the force $F(t)$ presents a strong periodic pattern. This pattern corresponds to runner's feet hits on the treadmill. The speeds oscillations in Figure 2 (Left) are due to the slowing down of the treadmill when the runner's feet hits it. Therefore, the runner should be considered to run at a constant speed.

In the Figure 2 (Right), the force applied on the treadmill $F(t)$ for each constant speed trial is displayed in a functional box-plot. The red curves are outliers on a band-depth point of view. Details on functional box-plot and the notion of band-depth can be found in Sun and Genton (2012). These outliers correspond to constant speed trials higher than $4.5m.s^{-1}$. This means that the trials at lower speed than $4.5m.s^{-1}$ forms an homogeneous

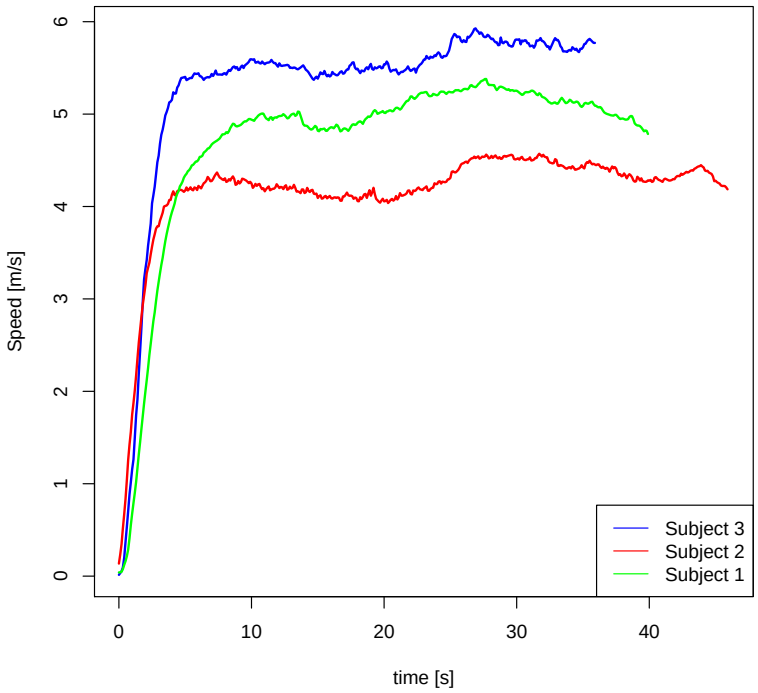

Figure 1: LMAM GPS measurement of instantaneous speed on outdoor 200 m runs. The sampling frequency is 10 Hz.

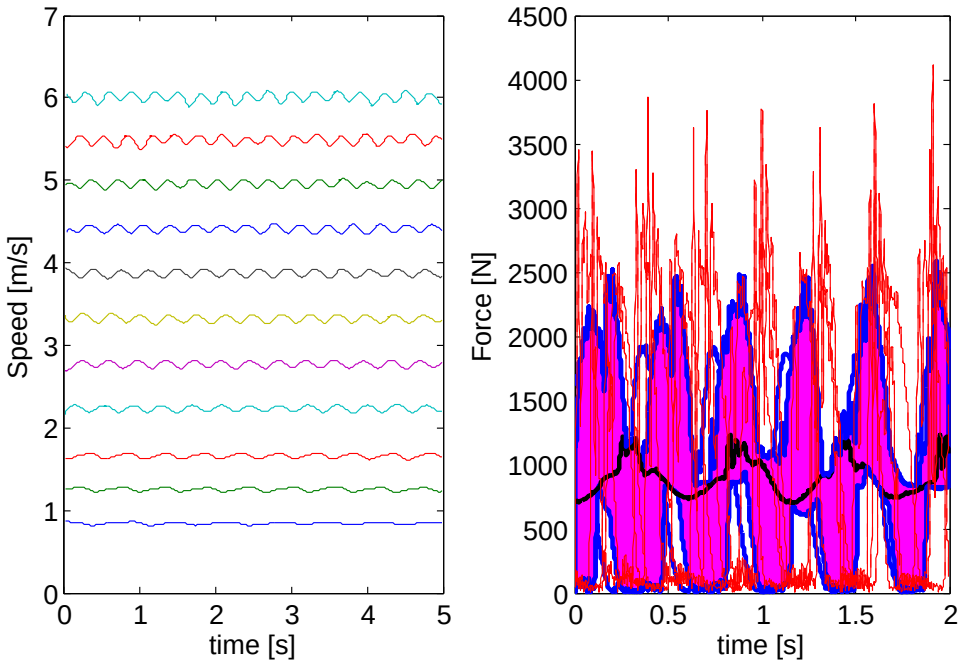

Figure 2: (Left) Speed on the treadmill for various zero-acceleration trials. (Right) Functional box-plot of the norm of the force $F(t)$ applied on the treadmill corresponding to the various trials.

group of curves. This group is represented by the purple zone delimited by the blue curves. While higher speed trials present a different force pattern in terms of magnitude, shape, or both. For larger speeds, the force applied on the floor $F(t)$ presents a higher frequency for the peaks occurrence. After 1 s the red curves present four peaks, while the group of homogeneous curves present only three. This reveals that another phenomenon takes place, and it is nothing else than the rise of runner's foot-pace. Indeed, the rise of foot pace desynchronizes signals. In addition, the value of the peaks is larger for larger speeds. Both phenomena contribute to the rise in propulsive force $f(t)$ and in speed $v(t)$. First the intensity of force applied on the floor, but also the rate at which this force is applied. Hence, it is difficult to exploit any information on $F(t)$ to extrapolate on $f(t)$ without a specific modelling.

From this preliminary analysis, we conclude that it is more appropriate to suppose a parametric form for $f(t)$. Indeed, even if some measurements of $F(t)$ were available along with the speed recordings of Figure 1, it would be difficult to exploit them.

In the following sections, the statistical context is made precise. In particular, the parameter's identifiability given the data available is discussed. Then, an overview of the reasons to introduce the specific methods to ODE parameter estimation are given.

2.2 Context for ODE's parameters identification

Suppose that for a collection of given times $\{t_i\} \in [0, T]$, one observes y_i:

$$y_i = v(t_i; \theta) + \epsilon_i, \quad i = 1, ..., N_{obs}.$$

The noise ϵ is supposed to be weakly stationary with $E(\epsilon) = 0$, $E(\epsilon^T \epsilon) = \Sigma$, and the velocity $v(t_i, \theta)$ satisfies the equation $\dot{v}(t; \theta) = F(t, v; \theta)$ for $t \in [0, T]$ associated to an initial condition. The parameter θ can be anything that makes the ODE problem well-defined: coefficients defining the homogeneous ODE, parameters defining the input function that is non-observed or even initial conditions. The ODE may not possess an explicit solution for non-linear function $F(t, v; \theta)$ in the second argument. In such context, an approximation $\hat{v}(t; \theta)$, satisfying $\dot{\hat{v}}(t; \theta) \approx F(t, \hat{v}; \theta)$, must be provided in order to exploit the non-linear model $v(\cdot; \theta)$. It is a well-known problem from numerical-analysis, and usually treated with a Runge-Kutta integration scheme. There are limitations on the use of numerical integration. First, it could be too costly: for each value for θ the numerical integration scheme must be repeated. In addition, it is expected to have a lot of updates for θ, because a non-linear model $\hat{v}(t; \theta)$ is fitted to the data y. Therefore, the optimization task becomes rapidly challenging. These limitations motivated the development of specific methods for parameter estimation in ODEs. In this report, we will study two of them. The general profiling method is introduced in Ramsay et al. (2007), and the two-step method is treated in detail in Brunel (2009). The general profiling method is closely related to the usual non-linear least squares one, while the two-step method is different. Before going into the method details, the identifiability of energetic parameters and the modelling of the propulsive force must be discussed.

2.3 Lack of identifiability for energetic parameters

In Section 2, we mentioned that $e(t)$ cannot be observed along a race. While this report explores inference methods for θ that tolerate unobserved state in the system of ODEs, the estimation problem of energetic parameters is more fundamental. Any parameters

related only to the energy equation, e.g., $\bar{\sigma}$ or e_0, are not identifiable ! In (1), the evolution equation for the velocity $v(t)$ does not depend on $e(t)$. For any values of $\bar{\sigma}$ and e_0, the velocity remains unchanged: $\frac{\partial v(t;\theta)}{\partial e_0} = \frac{\partial v(t;\theta)}{\partial \bar{\sigma}} = 0$. Therefore, updates neither for $\bar{\sigma}$ nor for e_0 bring the fitted velocity profile $\hat{v}(t;\theta)$ closer to the data y. The model fitting of speed profile will not lead to any information on e_0 or $\bar{\sigma}$. A simulation approach is the only alternative. For instance, from the fitted velocity profile/propulsive force $\hat{v}(t) = v(t;\hat{\theta})/\hat{f}(t) = f(t;\hat{\theta})$, and given values of e_0 and $\bar{\sigma}$, we could generate an explicit solution $\hat{e}(t;\bar{\sigma},e_0)$ of (1). If there is no uncertainty on e_0 and $\bar{\sigma}$, it is possible to quantify uncertainty on $e(t)$. While this assumption could be met for $\bar{\sigma}$, as it is a well-known quantity for runners, $\dot{V}O_2^{\max}$), the initial anaerobic energy e_0 remains inaccessible for runners.

2.4 Modelling of the propulsive force

Since the energetic parameters are not identifiable, the main part of the modelling concerns the propulsive force. The parameter θ reduces to runner's internal body friction τ and parameters related to the propulsive force $f(t;\theta)$. We will suppose a continuous piecewise linear function for $f(t;\theta)$ with N_{p} pieces:

$$f(t;\theta) = \sum_{k=1}^{N_{\mathrm{p}}} (\alpha_k t + \beta_k) \mathbb{1}[\eta_{k-1}, \eta_k](t), \quad t \in [0,T], \tag{3}$$

where α_k, β_k are the slope and intercept of each piece, and η_k the changing times satisfy $\eta_0 = 0 < \eta_1 < \cdots < \eta_{N_{\mathrm{p}}} = T$. The continuity requirement on the propulsive force maintains a physical meaning, and makes the values f_k at η_k for $k = 0, ..., N_{\mathrm{p}}$ more natural parameters than α_k and β_k. The changing times are unknown (except for $\eta_0 = 0$ and $\eta_{N_{\mathrm{p}}} = T$) and the parameter is $\theta = (\tau, f_0, ..., f_{\eta_{N_{\mathrm{p}}}}, \eta_1, ..., \eta_{N_{\mathrm{p}}-1})$. If one supposes N_{p} pieces for the force, then the dimension of the parameter space is $|\theta| = (N_{\mathrm{p}} - 1) + (N_{\mathrm{p}} + 1) + 1 = 2N_{\mathrm{p}} + 1$. The parametric form $f(t;\theta)$ may appear very restrictive. However, it should not have a drastic influence on the fitted curve $\hat{v}(t)$. One must recall that the ODE for the velocity is linear and has the explicit solution:

$$v(t;\theta) = \int_0^t \exp\left\{-\frac{(s-t)}{\tau}\right\} f(s;\theta) ds. \tag{4}$$

This implies that $\hat{v}(t)$ will be very smooth, as it is a convolution with an exponential. One could also argue that the path's regularity from $f_{\eta_{k-1}}$ to f_{η_k} is less relevant than an increasing/decreasing phase of the propulsive force $f(t;\theta)$. Besides, it simplifies the estimation task to avoid higher degree polynomial between changing times η_k.

2.5 Non-linear least squares method (NLS)

For a given parametric form $f(t;\theta)$, one obtains an explicit solution for $v(t;\theta)$ by computing the integral in (4). The first thing to try is to fit $v(t;\theta)$ to the data y by maximizing an appropriate log-likelihood:

$$\hat{\theta} = \underset{\theta \in \Theta}{\arg\max} \quad l_\Sigma(\theta), \tag{5}$$

where Σ is the noise covariance, and $l_\Sigma(\theta)$ is the log-likelihood for the model. For convenience, we will write $v(\theta) := v(\cdot;\theta)|_{\{t_i\}_{i=1}^{N_{obs}}}$, i.e., $v(\theta) \in \mathbb{R}^{N_{obs}}$. In the case of a

Gaussian behavior of the residual $\epsilon \sim \mathcal{N}_{N_{obs}}(0, \Sigma)$ the estimator becomes:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \quad \{y - v(\theta)\}^T \Sigma^{-1} \{y - v(\theta)\} - \log|\Sigma|. \tag{6}$$

It is usual to consider an ARMA structure for auto-correlation of ϵ in a non-linear regression context (see Chapter 6 in Seber and Wild (2005)). If the structure is correctly identified, then the standard asymptotic normality result applies. Under regularity conditions, this means that the distribution of $\hat{\theta}$ can be approximated by a normal law with a covariance estimated from the Fisher information $I(\theta)$. This asymptotic theory provides us with confidence interval and hypothesis testing for θ, in addition to likelihood ratio test to compare embedded models. For instance, it is useful to tests the need for additional pieces in the propulsive force $f(t;\theta)$, or the presence of significant trend α_k within a piece and so on. For practical purpose the fitting strategy is decomposed in three steps:

1. Compute $\hat{\theta}$ considering an i.i.d. noise.
2. From residuals of step 1. identify an ARMA structure and compute $\hat{\Sigma}$.
3. Re-estimate $\hat{\theta}$ and $\hat{\Sigma}$ using (6) and previous estimates as an initial guess.

Each of these steps may be challenging. Step 1 presents inherent difficulties to non-linear modelling that are related to numerical optimization. The optimization surface could present local optima, and an appropriate choice of initial guess is needed. Step 2 becomes problematic when the ARMA structure fails to model the residuals' correlation. Another correlation model could be considered but then Step 3 would be more difficult, although it is already a high-dimensional optimization problem.

2.6 General profiling method (GP)

This method is a relaxation of the non-linear least squares method. Instead of maximizing the likelihood among the exact solutions of the problem $\dot{v}(t;\theta) = F(t, v; \theta)$, the maximization is performed among relaxed solutions of this problem. The meaning of solutions in a relaxed-sense, and the procedure to compute them, will be explained in Section 3. For the time being, it is sufficient to acknowledge that a relaxed-solution lives in a ball of radius R (for a particular metric) around the exact solutions $v(t;\theta)$. From the study of Ramsay et al. (2007), one could distinguish two motivations to introduce this method:

- The optimization surface for the log-likelihood will present fewer local-optima for large radius (comparison between Fig.2 and Fig.5 in Ramsay et al. (2007) and Section 2.10). This behavior of the optimization surface for relaxed solutions suggests an iterative procedure to fit an ODE-based model. One begins with large R_0, then decreases it to R_1 using $\hat{\theta}^{R_0}$ as initial guess for R_1, i.e., $\theta_0^{R_1} = \hat{\theta}^{R_0}$, and so on until the stabilization of the estimate. This procedure is proposed in Qui and Zhao (2010) who conclude that as $R \to 0$, the estimate $\hat{\theta}^R$ should tend to the same one as in the NLS method, with a similar asymptotic distribution.

- The ODE model may not perfectly represent the underlying process. The flexibility of the choice of the neighborhood's radius R allows for model-mis-specification. Therefore, the ODE model is informative but not perfect. This interpretation of R is exploited to choose its correct value in Section 3.3.1. of Campbell and Chkrebtii (2013).

The first point defines a robust fitting procedure of an ODE model, in the sense that it should not be too sensitive to the initial guess, while it should possess asymptotic properties similar to the NLS estimator. For our estimation problem, a quick overview of the velocity profile of Figure 1 already roughly indicates the changing times, η_k, and physiological constraints also bring additional information. The second point is attractive for modelling of the propulsive force $f(t;\theta)$. Indeed, local departures of the speed profile general trend in Figure 1 may not present any interest for applications. This is likely to be even truer if the race is ten times longer than 200 m. In this framework, the GP method allows us to define a less complex ODE model while fitting the data equally well. The question reduces to investigating whether the estimate $\hat{\theta}$ remains meaningful with respect to the ODE model, and explores the validity of uncertainty quantification for mis-specified propulsive force. Some details on relaxed solutions are given on Section 3, along with the asymptotic theory for the GP estimator.

2.7 Two-step method

The previous two methods are similar, both are based the likelihood of the model $y = v(\theta) + \epsilon$. From Ramsay et al. (2007) and Qui and Zhao (2010), one could see the NLS method as a particular case of the GP method when the radius R tends to 0. On the other hand, the two-step method defines an M-estimator of θ from non-parametric estimator of velocity $\hat{v}$ and acceleration $\hat{\dot{v}}$. The first step consists of estimating $\hat{v}$, $\hat{\dot{v}}$ from usual non-parametric methods, e.g., kernel smoothing, spline smoothing, a local polynomial regression. Then an M-estimation criterion is defined as the distance $R(\theta)$ between $\hat{\dot{v}}$ and $F(t, \hat{v}; \theta)$. The motivation is again to diminish numerical optimization difficulties. The M-estimator criterion will tend to be easier to minimized because the non-linear behavior on θ tends to be avoided on the phase-plane $(v, \dot{v})$. Precisely, this means that the relation between v and $\dot{v}$ defined by $F(t, \hat{v}; \theta)$ tends to be linear in θ, or at least easily transformed to be linear. For our model, we will show in Section 4 that for given changing times η_k, the fitting procedure reduces to a linear model. The asymptotic theory for the two-step estimator is established in Brunel (2009). However, the covariance estimate inspired from asymptotic behavior behaves poorly. A bootstrap procedure might be an interesting idea, since the model-fitting is fast.

3 Generalized profiling estimation of parameters in ODE

3.1 Relaxed-solution of ODE for a given value of θ

Spline smoothing[1] and its more general version, the Reproducing Kernel Hilbert Space (RKHS), offer optimal approaches to non-parametric estimation. The Hilbert space containing the true underlying curve is denoted $\mathcal{H}$. In RKHS smoothing procedures a regularization parameter λ controls the extent to which the fitted curve should fall into a certain linear subspace of $\mathcal{H}$, usually defined as the kernel of a linear operator $L : \mathcal{H} \to \mathcal{H}$. For spline smoothing the operator is $L = D^2$, where $D^2 v = v''$, and the Hilbert space

[1]Spline smoothing will denote the use of a roughness penalty on the second derivative with knots at each data points. By opposition to regression spline smoothing that denotes a usual linear regression but on B-spline functions with user-defined knots.

is $\mathcal{H} = H^2([0,T]) = \left\{\mu : [0,T] \to \mathbb{R} : \mu, \mu' \text{ are absolutely continuous and } \int_0^T \{\mu''(t)\}^2 \, dt\right\}$. In this context increasing the value of λ restricts the class of fitted curve to linear functions, since $v \in \text{Ker}(L) = \text{Span}\{1, t\}$. The other way around, for decreasing λ, the fitted curves get closer to interpolations of the data. The properties of this smoothing procedure are remarkable. Firstly, the fitted curve $\hat{v}$ is the closest function in $H^2([0,T])$ to the data in a least-squares sense under the constraint that $||L\hat{v}||_{L^2} < r(\lambda)$. Secondly, the computation of $\hat{v}$ can be performed on a finite-dimensional space while the minimization space $H^2([0,T])$ is infinite-dimensional. In Heckman (2011) the author presents the theory and application of RKHS in great simplicity. In particular, Section 6 of Heckman (2011) treats the case of a linear differential operator of the form $Lv = v^{(m)} + \sum_{j=0}^{m-1} w_j v^{(j)}$ with w_j a real-valued and continuous function.

An ODE problem with a given value of θ can be written as follows: find $v_\theta \in \Gamma$ s-t $L_\theta v = f_\theta$, where f_θ is the forcing term and Γ a functional space that makes the operator L_θ well-defined. In our case $L_\theta = D + \tau^{-1} I$ and it is a linear operator on $H^1([0,T])$. This re-writing suggests a close link with the previous discussion. The main difference comes from the dependence on θ. Essentially, we would like the estimator for $v(t,\theta)$ to be a relaxed solution of the ODE problem while fitting the data well. A minimization criterion that generates such an estimator is of the form:

$$J(v \mid \theta, \lambda) = -\log\left[g_\Sigma\left\{y - v(\cdot;\theta)|_{\{t_i\}_{i=1}^{N_{obs}}}\right\}\right] + \lambda \int_0^T \{L_\theta v(t) - f(t;\theta)\}^2 \, dt \tag{7}$$

where g_Σ is the density of the error vector ϵ, y the vector of observations and λ a positive real-valued number. The minimizer of (7) will be denoted $\hat{v}(t;\theta,\lambda)$. From the Lagrange multipliers property, this estimator solves the equivalent problem:

$$\hat{v}(t;\theta,\lambda) = \underset{v \in H^1([0,T])}{\arg\min} \; -\log\left[g_\Sigma\left\{y - v(\cdot;\theta)|_{\{t_i\}_{i=1}^{N_{obs}}}\right\}\right],$$

$$\text{under the constraints that } ||L_\theta v - f_\theta||_{L^2} < r(\lambda),$$

where $r(\lambda) \to 0$ as $\lambda \to \infty$. From this equivalent formulation, it is clear that the minimization of $J(\cdot \mid \theta, \lambda)$ leads to an estimator that is a relaxed solution of the ODE problem. For a given θ, the ODE model relaxation allows to bring the the estimator $\hat{v}(t;\theta,\lambda)$ closer to data. In Section 2.6, we motivated the introduction of the GP method with a certain neighborhood R of ODE solutions. In the present section, the neighborhood R corresponds to $r(\lambda)$. It gives an interpretation to λ, i.e., it controls, in L^2-norm, to which extent $\hat{v}(t;\theta,\lambda)$ fails to satisfy the ODE $L_\theta v = f_\theta$.

The problem reduces to finding a practical solution to the minimization of (7) over an infinite-dimensional space $H^1([0,T])$. The theory of RKHS does not provide a finite-dimensional equivalent of this problem because $\text{Ker}(L_\theta(\cdot) - f_\theta)$ is not a linear subspace of $H^1([0,T])$, because $0 \notin \text{Ker}\{L_\theta - f_\theta\}$. Even if the ODE is homogeneous, i.e. $f_\theta = 0$, the operator L_θ may not be linear, and then the RKHS procedure also fails. A tractable approximation for the minimization problem of (7) is suggested in Ramsay et al. (2007). It relies on a collocation approximation, i.e., restricting the minimization space to finite dimension. It is very common to use a space spanned by B-spline basis functions $\{\phi_i\}_{i=1}^{N_{basis}}$ of order m over a set of knots associated to sampling times $K = \{\kappa_0 = 0, \kappa_1 = t_1, ..., \kappa_{N_{obs}} = T\}$. The dimension of the space spanned is $N_{basis} = (|K| - 2) + m = N_{obs} - 2 + m$, and will be denoted $S_K^m = \text{Span}\{\phi_1, ..., \phi_{N_{basis}}\}$. The approximation to the estimator $\hat{v}(t;\theta,\lambda)$ is defined by

$$\tilde{v}(\cdot;\lambda,\theta) = \underset{v \in S_K^m \cap H^1([0,T])}{\arg\min} \; J(v|\theta,\lambda). \tag{8}$$

Then, it is possible re-write this problem as a finite-dimensional optimization, with N_{basis} unknowns, through the formula $v(t) = \sum_{i=1}^{N_{basis}} \phi_i(t)c_i = c^T\Phi(t)$:

$$\hat{c}(\theta, \lambda) = \underset{c \in \mathbb{R}^{N_{basis}}}{\arg\min} J(c^T\Phi(\cdot)|\theta, \lambda).$$

Considering the criterion $J(c^T\Phi(\cdot)|\theta, \lambda)$ in its finite-dimensional form, one should note that λ acts as a regularization parameter. If $\lambda = 0$, a solution is given by an interpolation of the data, and for $\lambda \to \infty$, the solution is given by a best-fitting approximation of ODE solutions with functions in S_K^m. The latter statement is not (at all) straightforward, but one can find a proof in Qui and Zhao (2010) (Theorem 3.1). For the infinite-dimensional problem, it is easier to be convinced that as λ goes to $+\infty$, $\hat{v}(\cdot, \theta, \lambda)$ tends to the closest fit to data among ODE solutions $v(\cdot, \theta)$ (see Section 2.9 in Ramsay et al. (2007)). The distance between the best approximation functions in $S_{K_n}^m$ and ODE solutions is denoted by d_n. K_n is a collection of knots that increases with n the number of sampling points. Theorem 3.1 in Qui and Zhao (2010) states that if λ_n is an increasing sequence that tends to $+\infty$ while d_n tends to 0, then:

$$\sup_{\theta \in \Theta} ||\tilde{v}(\cdot, \lambda, \theta) - v(\cdot, \theta)||_{L^\infty} \leq [O_p(\lambda_n^{-0.5}) + d_n]C_T \tag{9}$$

where C_T is a constant independent of θ and n. This is an interesting result because it shows that the L^∞ error between ODE solutions and the collocation approximation is a sum of two independent contributions. The first is related to λ and the stochastic properties of the data. The second is deterministic and relies only on approximation properties of S_K^m. Such a contribution would not exist for the infinite-dimensional estimator $\hat{v}(\cdot, \theta, \lambda)$. From this result it is clear that for large enough λ, it is pointless to increase its value more as d_n dominates the error. The only improvement possible is to tailor the choice of the collocation space S_K^m so that d_n is as small as possible. This is equivalent to choose a basis system and it is an issue for any collocation method. The choice of the basis system can be made optimal for homogeneous and linear ODE with RKHS theory, and it is an optimal choice in the sense that $d_n = 0$ for any n. This optimal choice of basis is an another way to state that we solve an infinite-dimensional problem with the help of an equivalent finite-dimensional problem. This idea has not been discussed[2] since the GP method is particularly attractive for non-linear problem where explicit solutions are not available. In the following, the distinction between $\tilde{v}(\cdot, \lambda, \theta)$ and $\hat{v}(\cdot, \lambda, \theta)$ will be ignored, and we will denoted $\hat{v}(\cdot, \lambda, \theta)$ as the solution of (8).

3.2 Parametric estimation of θ

The previous section shows how to produce a relaxed solution $\hat{v}(\cdot, \lambda, \theta)$ of the ODE problem, well-defined for a given value of θ. The departure from the ODE model is monitored via λ, considered as fixed in this section. The criterion defining the relaxed solution is:

$$H(\theta) = -\log\left\{g_\Sigma(y - \hat{v}(.; \theta, \lambda)|_{\{t\}_i})\right\}. \tag{10}$$

Along with the minimization of $H(\theta)$, one needs to update the relaxed solution $\hat{v}(\cdot, \lambda, \theta)$. From an initial guess θ_0 and a fixed value for λ:

1. compute the minimizer of (8) for θ_0 : $\hat{v}(t; \theta_0) = \sum_{i=1}^{N_{basis}} \hat{c}_i(\theta_0)\phi_i$;

[2] As far as the author is aware of.

2. update the guess using the gradient of $H(\theta)$;

3. return to 1. until $H(\theta)$ is minimized;

4. define $\hat{\theta}$ as the last guess.

This latter procedure is similar to the NLS algorithm for a relaxed solution $\hat{v}(\cdot, \lambda, \theta)$. In general, one would need to do an optimization to obtain the relaxed solution at each step. For our linear ODE of the velocity, an explicit solution for $\hat{v}(\cdot, \lambda, \theta)$ is available. The derivation is described in Appendix A.1, and the optimization procedure in Section 5.2. For non-linear ODE, this method remains computationally attractive because it is usually cheaper to perform the optimization of step 1 than to use a Runge-Kutta integration scheme at each step. In addition, the GP method can be applied for an increasing sequence of values for λ. This approach could avoid local minima that may arise for a large value of λ since a non-linear model is fitted to data (using the procedure in Section 2.6). This is because it is expected that the optimization surface $H(\theta)$ becomes more convex when λ is small. This latter statement is by no means guarantee in theory but it is a conjecture from Section 2.8 in Ramsay et al. (2007).

For the present model, these computational advantages are not essential since explicit solutions for the ODE exist. However, the propulsive force $f(t; \theta)$ has an unknown structure. The framework of relaxed solutions may avoid over-parameterization of $f(t; \theta)$, and avoid increasingly complex modelling as the race gets longer. In addition, for a longer race, local departures of the global trend of the speed profile may present no interest. The asymptotic distribution of the GP estimator $\hat{\theta}(\lambda)$ is established in Qui and Zhao (2010). This asymptotic theory controls the estimator's behavior as $\lambda \to +\infty$. While the proof is complicated, the main idea behind it is fairly simple. From the approximation result for relaxed solutions of (9), one could expect the behavior of the GP estimator to be similar to that of the NLS estimator as $N_{obs} \to +\infty$. The approximation error $\tilde{v}(\cdot, \lambda, \theta) \approx \hat{v}(\cdot, \lambda, \theta)$ will tend to 0 as the collocation space S_K^m gets richer and the radius $r(\lambda)$ around the ODE solutions becomes smaller and smaller. Qui and Zhao (2010) describe the simultaneous control of error from spline approximation and relaxed solutions. The main result states that for a large λ and a large number of observations:

$$\hat{\theta} \overset{d}{\approx} \mathcal{N}(0, I(\theta)^{-1}),$$

where $I(\theta)$ is the expected Fisher Information. Thus, the GP estimator is asymptotically efficient. Even if it is not guaranteed by the theory, in the context of a middle to low value for λ, the same variance estimator is used. After replacing the expected Fisher Information by its estimator counter-part, the covariance estimator for GP method is:

$$\mathrm{Cov}(\hat{\theta}(\lambda)) \approx \left(\frac{d^2 H}{d\theta^2}(\theta; \lambda)|_{\theta=\hat{\theta}} \right)^{-1}. \tag{11}$$

This idea was proposed in Section 3.2.1 of Campbell and Chkrebtii (2013). Its rationale comes from the dependence on λ of the optimization surface $H(\theta; \lambda)$. For small λ the surface should become flatter. This may account for additional variability in the parameters when the model is largely relaxed. The covariance estimator, defined on 11, is used in Section 5.2. Another covariance estimator, proposed in Section 2.9 of Ramsay et al. (2007), is based on a linearization of $\hat{\theta}(y)$, and is summarized in Appendix A.2. Therefore, its quality is difficult to assess in general, as this Taylor expansion may or not be an accurate

$\log\lambda$	1.5	2.0	2.5	3.5	4.5	5.0	NLS
Bias τ	0.517	0.349	0.187	0.026	0.003	0.003	0.006
Bias $f_{\max}$	−2.938	−2.290	−1.316	−0.106	0.089	0.059	0.059
$SD_{\text{simu}}(\hat{\tau})(\lambda)$	0.17	0.16	0.13	0.12	0.12	0.11	0.11
$SD_{\text{simu}}(\hat{f}_{\max})(\lambda)$	0.67	0.81	0.87	1.05	1.15	1.08	1.05

Table 1: Summary statistics of GP and NLS methods for simulated data.

approximation. For the simulation study of next section, this covariance estimator is used because it is directly available in the `CollocInfer` package. Its bad performance suggests employing the estimator of (11), inspired from asymptotic theory.

3.3 Simulation results for the GP method

In this section the GP estimator $\hat{\theta}(\lambda)$ is compared to the NLS estimator. The data are generated by adding Gaussian i.i.d. noise with $\sigma = 0.5$ to the analytic solution in the sprint case, i.e., $f(t;\theta) = f_{\max}$ and hence $v(t) = f_{\max}\tau\left\{1 - \exp(-\frac{t}{\tau})\right\}$. The parameters are $\tau = 1$ and $f_{\max} = 9$. Only the case of 100 observations is studied. The collocation approximation is performed on the space S_K^4, that is spanned by a cubic spline basis with knots associated to each observation. The fitting procedure is performed with `Profile.LS()` function from the `CollocInfer` package.

In Figure 3 (Top), the bias $\left\{\hat{\theta}(\lambda)\right\}$ is a decreasing function of the regularization parameter λ. This is expected since the underlying process exactly follows the ODE model, and the GP method for large λ tends to have behavior similar as NLS method. This latter is unbiased, and hence bias$(\hat{\theta})$ should decrease with λ. The sampling pattern of the NLS estimator in Figure 4 mimics the GP estimator for $\log\lambda \geq 4.5$ in Figure 3 (Top). This observation is confirmed in Table 1. It is remarkable that the basis approximation of the GP method does not deteriorate the sampling variability of resulting estimator $\hat{\theta}$. The approximation quality stops improving for $\log\lambda \geq 4.5$, as the estimator sampling pattern in Figure 3 (Top) remains identical for larger values.

Figure 3 (Middle) shows that the variance estimator based on linearization argument. It over-estimates the variability for low λ, and under-estimates it for large λ. For this reason the covariance estimator of (11) is preferred for real-data fitting in Section 5.2. The sampling variability does not clearly decrease with λ: while the ODE model is relaxed, the fitting criterion $H(\theta;\lambda)$ does not account for this mis-specification. This issue is further explored on the real data.

In Figure 3 (Bottom), the integrated bias $E\left[\int_0^T \left\{\hat{v}(t;\hat{\theta}) - v^*(t;\theta^*)\right\} dt\right]$, and the integrated MSE $E\left[\int_0^T \left\{\hat{v}(t;\hat{\theta}) - v^*(t;\theta^*)\right\}^2 dt\right]$ decreases with λ to a positive limit, hence the integrated variance decreases as well to a positive limit. This is expected from (9) because the basis approximation error d_n dominates the error for large λ: a finite number of cubic spline functions cannot arbitrarily well approach the exact solution of the ODE. The saturation of the approximation error also explains the sampling pattern with respect to λ of Figure 3 (Top).

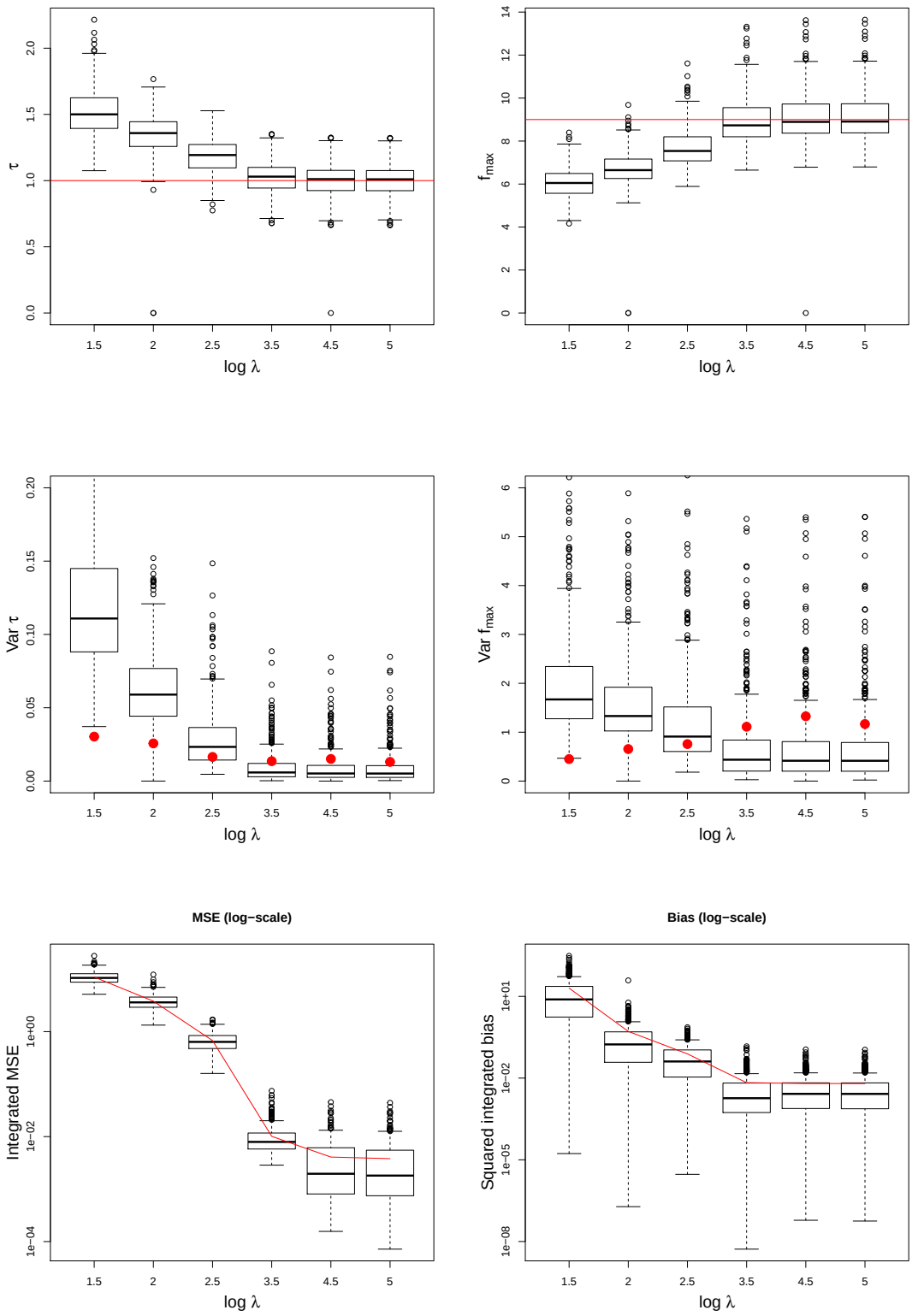

Figure 3: GP method diagnostics for $\hat{\theta}(\lambda)$ and $\hat{v}(\hat{\theta}, t)$. 500 replications of the sprint model for $N_{obs} = 100$ with i.i.d. Gaussian noise and $\sigma = 0.5$. (Top) $\hat{\tau}(\lambda)$ and $\hat{f}_{\max}(\lambda)$ with red line representing true values. (Middle) Estimated variance computed with the linearization argument as described in (24), red dot represents observed sampling variance. (Bottom) Realisations of integrated bias and MSE, the red line is the mean of these quantities among simulations.

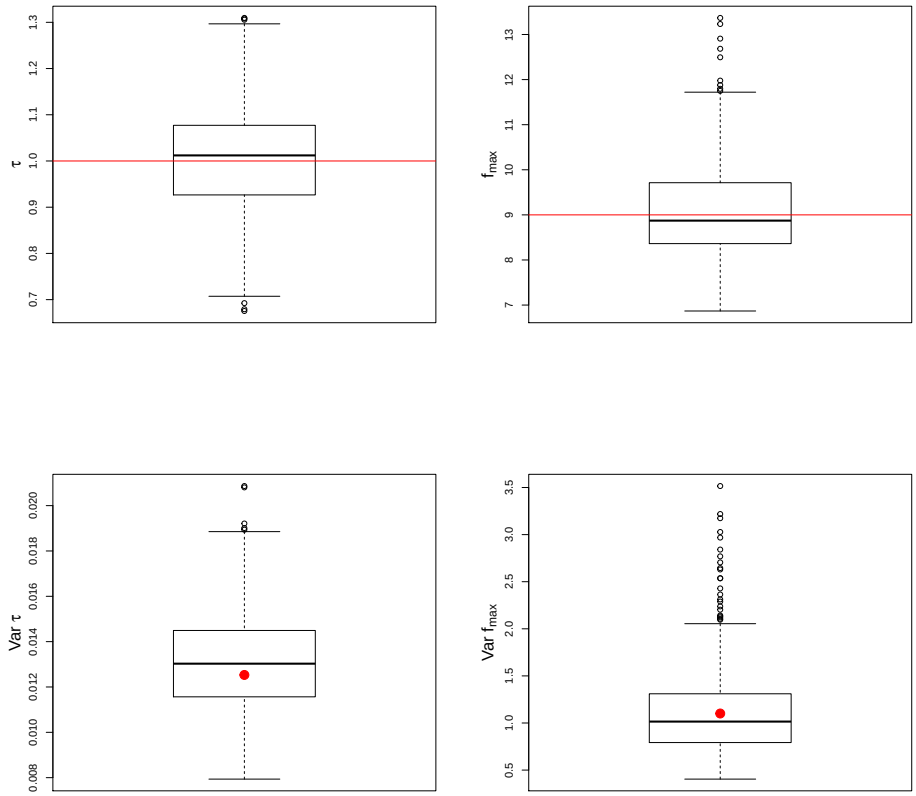

Figure 4: Same noisy data than in Figure 3. Diagnostics for non-linear least squares: (Top) $\hat{\tau}$ and $\hat{f}_{\max}$ with red line representing true values. (Bottom) Estimated variance using usual non-linear least squares asympotic distribution.

4 The two-step method

4.1 Principle

The two-step method exploits the ODE model formulation $\dot{v}(t) = F\{t, v(t), \theta\}$ to define an M-estimator criterion. The estimator for θ is defined through the minimization of:

$$\hat{R}^2(\theta) = ||\hat{v} - F(\cdot, \hat{v}; \theta)||_{L^2} = \int_0^T \left\{\hat{v}(s) - F(s, \hat{v}(s); \theta)\right\}^2 ds, \tag{12}$$

where $\hat{v}$ and $\hat{v}$ are non-parametric estimators of the functions $\dot{v}$ and v that satisfy the ODE model $\dot{v}(t) = F\{t, v(t), \theta\}$ for $0 < t < T$. Any other distance $R(\theta)^q = ||\hat{v} - f(., \hat{v}; \theta)||_{L^q}$ measuring the departure of the non-parametric estimator from the ODE model could be considered. More generally, it is useful to introduce a weight function $w(t)$ in the discrepancy function to account for non-parametric estimator's behavior locally: $R_w^q(\theta) = \int_0^T (\hat{\dot{v}}(s) - F(s, \hat{v}(s); \theta))^2 w(t) ds$. In particular, it is well known that kernel smoothers do not behave well at the boundaries of the time domain. Thus, downweigthing the impact of the non-parametric estimator at the boundary is important. For instance, Theorem 3.1 in Brunel (2009) demonstrates that for a weight function $w(t)$ that doesn't vanish at the boundary, the $\sqrt{n}$-consistency of estimators $\hat{\theta}$ is replaced by a slower order of convergence

$O(\sqrt{n|\xi_n|})$, where $|\xi_n|$ is the number of knots of the cubic spline basis used, and it must satisfy $|\xi_n|\sqrt{n} \to 0$ and $|\xi_n| n \to +\infty$.

The first step consists of constructing a consistent estimate $\hat{v}(t)$ and $\hat{\dot{v}}(t)$ from noisy observations $(t_i, v_i)_{i=1}^{N_{obs}}$. This smoothing problem has been widely treated in literature, in theory, all of these methods provide consistent estimates of the underlying curve v^*, supposed in our context to satisfy the ODE model $\dot{v}^*(t) = F(t, v^*, \theta^*)$.

The second step defines $\hat{\theta}$ from these non-parametric estimators and the discrepancy function R_w^q:

$$\hat{\theta} = \arg\min_{\theta \in \Theta} R_w^q(\theta).$$

This estimator doesn't explicitly involve the data, which is only exploited during the first step. In Brunel (2009), one can find a precise analysis of the asymptotic behavior of the estimator $\hat{\theta}$. Under fairly general conditions this paper demonstrates consistency for a general discrepancy function R_w^q, and asymptotic normality in the particular case of $q = 2$ and a B-spline basis for the smoothing part. The hypotheses for consistency of $\hat{\theta}$ are: the consistency of $\hat{v}(t)$, $\hat{\dot{v}}(t)$ and an identifiability condition on the asymptotic criterion R_w^q. The additional hypothesis concerns regularity of the function defining the ODE model $F(t, v(t); \theta)$, and its derivatives.

The reason why this method was introduced is because it doesn't rely on an explicit solution of the ODE. From a geometric point of view, the second step of the method is a functional regression in the phase-plane defined by $(v, \dot{v})$. Considering the infinite-dimensional output $Y(t) = \hat{\dot{v}}(t)$ and the infinite-dimensional explanatory variable $X(t) = \hat{v}(t)$, we are looking for parameters θ so that in a L^2-sense, $Y(t)$ is closest to $F(t, X(t), \theta)$. The relation between $X(t)$ and $Y(t)$, defined by $F(t, \cdot, \theta)$, is known up to the parameter θ, and this relation is usually much simpler than the equation linking θ and t to $v(t; \theta)$. Even for a linear ODE, the relation between $v(t; \theta)$ can become non-linear with respect to θ. For instance in the sprint model $f(t; \theta) = f_{\max}$:

$$\dot{v}(t) = f_{\max} - v(t)\tau^{-1}, \quad v(0) = 0 \implies v(t; \theta) = f_{\max}\tau \left\{1 - \exp(-\frac{t}{\tau})\right\} \tag{13}$$

the equation linking τ and $f_{\max}$ to the speed v is non-linear. To fit such a model to the speed data one needs to use the `nls()` command. While in the phase-plane, if one has a non-parametric estimator for the speed $\hat{v}$ and the acceleration $\hat{\dot{v}}$, the problem of estimating τ and $f_{\max}$ reduces to estimate an intercept $\hat{f}_{\max}$ and a slope $1/\hat{\tau}$. That is to say that $\hat{f}_{\max}$ and $\hat{\tau}$ are such that $||\hat{\dot{v}} + (1/\hat{\tau})\hat{v} - f_{\max}||_{L^2}$ is minimized. The L^2 distance is approximated using a quadrature formula:

$$\begin{aligned} \hat{R}_w^2(\theta_1, \theta_2) &= \int_0^T \left\{\hat{\dot{v}}(s) + v(s)\theta_1 - \theta_2\right\}^2 w(s) ds \\ &\approx \sum_j \Delta_j w(t_j) \left\{\hat{\dot{v}}(t_j) + v(t_j)\theta_1 - \theta_2\right\}^2 \end{aligned} \tag{14}$$

where $\theta_1 = 1/\tau$ and $\theta_2 = f_{\max}$. Thus the estimator $\hat{\theta}$ can be computed using the `lm()` command and with some appropriately chosen weights Δ_j and mesh points t_j such that the quadrature formula is close to the integral. The uncertainty for $\hat{\theta}$ cannot be computed using `lm()` standard errors because the linear model's covariance estimator does not account for the errors introduced in the first step by the non-parametric estimators $\hat{v}$ and $\hat{\dot{v}}$.

4.2 Simulation results for the two-step method

4.2.1 Computations & Simulation details

In order to better understand the properties of the two-step estimator with respect to the number of observations N, and the weight function $w(t)$, we explore the sprint problem. The noisy data are simulated using i.i.d. normal noise with $\sigma = 0.5$ added to $v(t) = f_{\max}\tau(1 - \exp(-\frac{t}{\tau}))$ for $t \in [0, 100]$. We used B-spline basis for the smoothing with a fixed number of basis functions $N_{basis} = 100$ and $N = 50, 100, 200, 1000, 2000, 10000$. The weight function equals the identity. In Figures 5 & 6, we present simulation results for 100 realizations of the noise. In Figure 5, the box-plot of estimates tends to show that as N increases, the bias and the variance of estimators vanishes, hence the MSE decreases as well. In Figure 6, this behavior is expected when one looks to a phase-plane plot of the non-parametric estimator $(\hat{v}, \hat{\dot{v}})$. As N increases the estimated curves on the phase-plane plot get closer to the straight line pattern that you find for the true solution ODE. Each of the estimates $(\hat{\tau}, \hat{f}_{\max})$ is produced using only one of the curves shown in Figure 6. The two following points summarize the computational details that led to $(\hat{\tau}, \hat{f}_{\max})$ for this example:

- For a basis of order-4 splines with equi-spaced knots with $N_{basis} = 100$, we perform a linear regression (that is regularized for the case $N = 50, 100$ with $\lambda = 0.1$) to obtain $\hat{v}(t) = \sum_{i=1}^{N_{basis}} \hat{c}_i \phi_i(t)$ and $\hat{\dot{v}}(t) = \sum_{i=1}^{N_{basis}} \hat{c}_i \dot{\phi}_i(t)$. This smoothing process is directly available with the `smooth.basis()` command of the `fda` package.

- From these non-parametric estimators, we define a linear model fit w.r.t. to a mesh $P = \{t_0 = 0 < t_1 < \cdots < t_{N_{mesh}} = T\}$ to approximate the integral defining penalty $\hat{R}_w^2$ as described in (14). This procedure corresponds to define $Y = \hat{\dot{v}}|_P \in \mathbb{R}^{N_{mesh}}$ and $X = \hat{v}|_P \in \mathbb{R}^{N_{mesh}}$, and performs the weighted least squares projection of Y onto Span$\{1, X\}$. The weights are defined by $W_j = \Delta_j w(t_j)$ $j = 0, 1...N_{mesh}$, where Δ_j are the weights of the quadrature formula chosen (trapezoidal rule here) and $w(t)$ the penalty weight function.

In order to the previous steps to become feasible, even for $N_{simu} = 100$, it is crucial to vectorize the computation. For the smoothing part, the `smooth.basis()` command deals with it easily: from a matrix of observations $\mathbb{R}^{N_{obs} \times N_{simu}}$ the function outputs a matrix of coefficients $\mathbb{R}^{N_{basis} \times N_{simu}}$. For the linear model, we construct augmented output $\tilde{Y} = (Y_1^T, ..., Y_{N_{Imus}}^T)$ with $Y_i \in \mathbb{R}^{N_{mesh}}$ and an augmented design matrix $\tilde{X}$ as a block diagonal matrix with blocks corresponding to $[1, X_i] \in \mathbb{R}^{N_{mesh} \times 2}$. As a consequence, all the simulation estimates are computed with one call to `lm()`. This vectorization is particularly needed since we want to approach an integral criterion. Thus, N_{mesh} might be big.

4.2.2 Comment on results

Figure 5 testifies that the two-steps method is consistent. Indeed, the bias and the variance for $(\hat{\tau}, \hat{f}_{\max})$ tend to zero in Figure 5, hence the MSE tends to zero as well which implies the consistency. However, even for $N = 10000$ the distribution around true values is a little bit biased. It must be recall that the LMAM's GPS equipment is only able to measure the speed with 10 Hz frequency. Thus, it is important to understand how to improve the two-step method. Figure 6 gives us insight on what is going wrong. Indeed, the known bad behavior of the non-parametric estimator near the origin is clearly

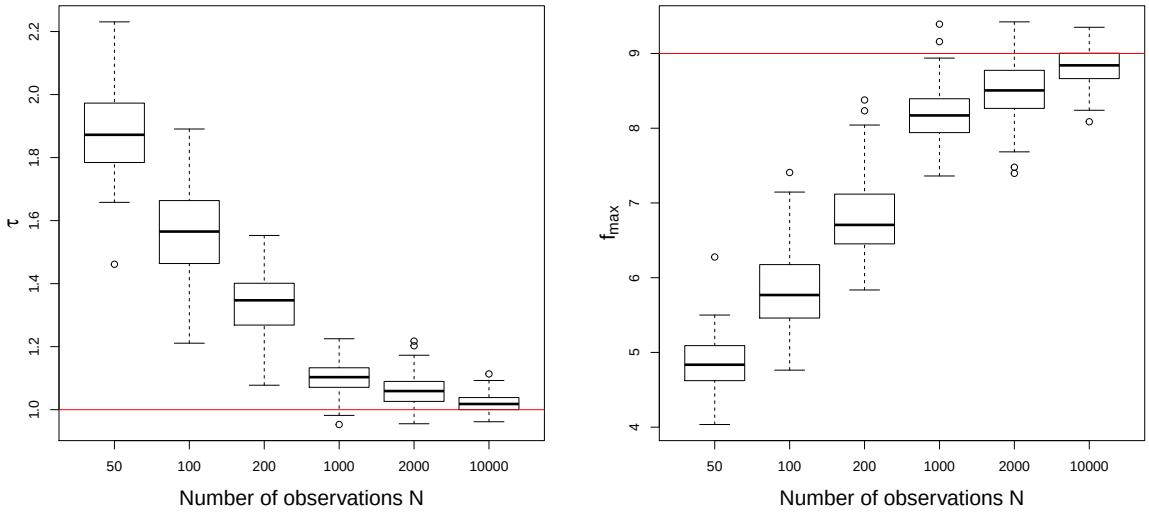

Figure 5: Histogram of estimates for 100 simulations of the sprint model, with i.i.d. Gaussian noise for $\sigma = 0.5$ for increasing N. The red lines indicate true values of τ and $f_{\max}$.

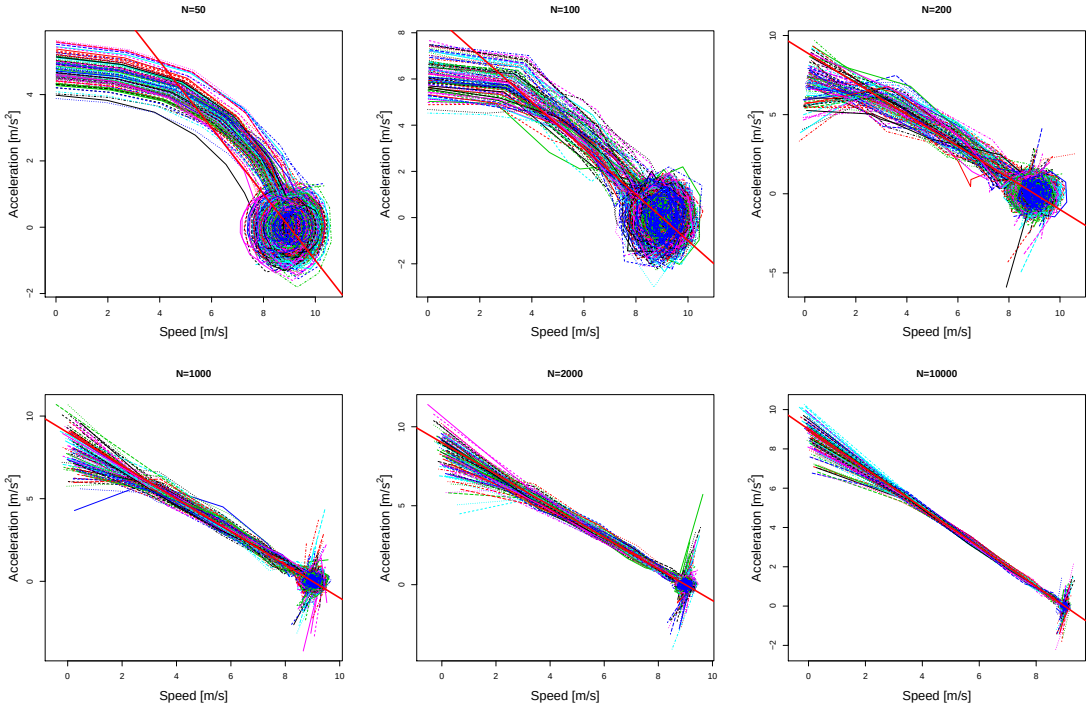

Figure 6: Phase-plane plot of the non-parametric estimators $\hat{v}$ and $\hat{\dot{v}}$ associated to the 100 simulations of the sprint model, for increasing N. The red line indicates the real phase-plane plot $\dot{v} = -v + 9$.

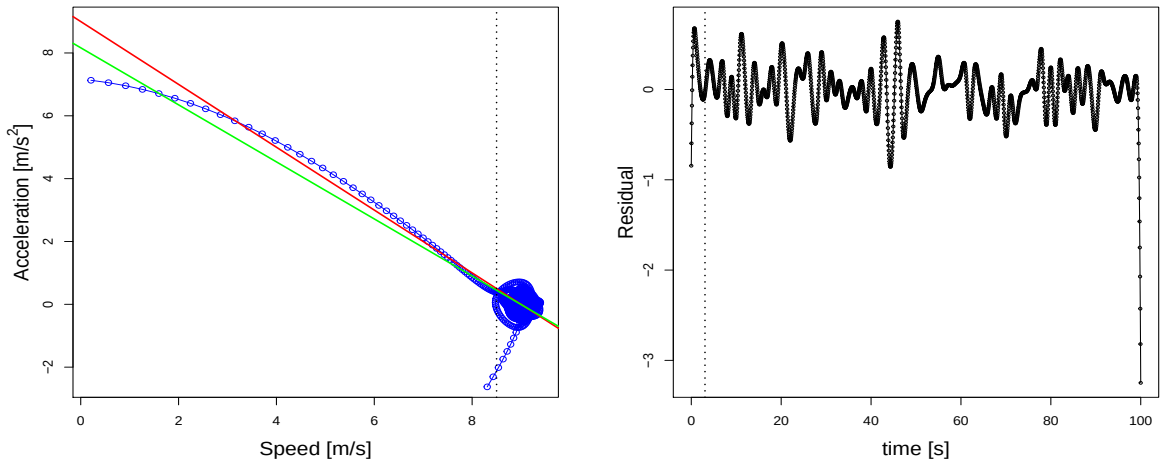

Figure 7: A typical regression for N=1000. (Left) Blue points indicate the evaluation of $\hat{v}$,$\hat{\dot{v}}$ on the mesh. Green line is the fitted regresion and Red line is the true slope. (Right) Residual of the fit. The dot vertical line indicates the beginning of spaghetti-bowls.

apparent. The acceleration is underestimated for t near 0 until N goes bigger than 1000 observations. Even for a larger N, the curves present a skewed behavior around the red line. In addition, we observe a "spaghetti-bowls" around the zero acceleration's region (bottom-right). This is due to the wiggling nature of the smoothing step, causing the derivatives of the estimated curve to alternate between negative and positive values.

These two problems are illustrated in Figure 7 (Left) for a particular simulation with $N = 1000$ observations. Points with speeds below 3.5 are quite influential on the green-fitted curve, as well as the last observations (corresponding to almost orthogonal points to the fit). Even though the "spaghetti-bowls" looks bad, it is not of a great influence on fitted values since the derivative is oscillating around zero. In the residual of Figure 7 (Right), there is a clear correlation pattern mainly due to these oscillations of the smoothing spline used to estimate the velocity and acceleration profile. This latter observation confirms that it is really inappropriate to quantify uncertainty on $\hat{\theta}$ using the `lm()` standard errors.

To achieve faster convergence for $(\hat{\tau}, \hat{f}_{\max})$, we need to appropriately discard observations. For instance, we could use the weight function $w(t)$ decaying for t near 0 or 100. Perhaps adding knots during the increasing speed phase will improve cubic spline smoother performance for the acceleration $\hat{v}$ near origin. This idea necessitates a particular design, while playing with the weight function comes naturally as a control feature of the two-step method. In addition, it doesn't seem that we need much information to fit that straight line, so even discarding observations with weights is not a issue. Another option would be to change the loss function into a more robust one. For instance, choosing q=1 in the criterion R_w^q corresponds to a L^1 penalization.

5 Real data fitting

This section compares the performance of the NLS, the GP and the two-step methods based on the data displayed on Figure 1. The comparative analysis is presented for Subject 3. For the purpose of comparison, the ODE model for $v(t)$ is interesting because

an explicit solution is available for any propulsive force $f(t;\theta)$ defined on (3). The NLS method exploits this explicit solution, whereas the GP and the two-step methods use the ODE formulation defined by the propulsive force. Using the notation of (3), the explicit formula for the velocity for $t \in (\eta_{k-1}, \eta_k]$ is:

$$v(t) = e^{-(t-\eta_{k-1})/\tau} v(\eta_{k-1}) + \beta_k \tau [1 - e^{-(t-\eta_{k-1})/\tau}] + \alpha_k \tau [t - \tau + (\tau - \eta_{k-1}) e^{-(t-\eta_{k-1})/\tau}] \quad (15)$$

where $k = 1, ..., N_{\mathrm{p}}$, $\eta_0 = 0$, $\eta_{N_{\mathrm{p}}} = T$ and $v(0) = 0$.

5.1 NLS method with ARMA residuals

Following the method described in Section 2.5, we suppose the data y_t follow the model:

$$y_t = v_t(\theta) + \epsilon_t, \quad t = 1, ..., N_{obs}, \quad (16)$$

where $v_t(\theta)$ is an analytic solution defined by (15), and ϵ_t is an error process, supposed to be a Gaussian ARMA(p, q), and AR/MA parameters are noted respectively ϕ and κ. For a fixed number of changes, N_{p}, in the propulsive force $f(t;\theta)$ the parameter vector is $\theta = (\tau, f_0, f_1, ..., f_{\eta_{N_{\mathrm{p}}}}, \eta_1, ..., \eta_{N_{\mathrm{p}}})$. In the fitting process, the number of changes N_{p} must be fixed first. A crude guess can be derived from a graphical inspection of Figure 1. From this guess of N_{p}, we may consider that the model needs to be more/less complex and change N_{p} accordingly. Of course such a decision should be supported by statistical testing. The likelihood ratio test allows to compare the most complex model, for a given N_{p}, to a sequence of simpler models. In particular, we will use the likelihood ratio test to determine if within each piece $t \in (\eta_{k-1}, \eta_k]$, the forcing term $f(t;\theta)$ presents a significant trend α_k. In order to exploit the likelihood ratio test, we need to identify an appropriate correlation structure Σ in the data. If it is the case then usual asymptotic results on $\hat{\theta}$ would apply. In such a case, the approximate distribution for $\hat{\theta}$ is:

$$\hat{\theta} \overset{d}{\approx} \mathcal{N}(\theta, \widehat{I(\theta)}^{-1}) \quad (17)$$

where $I(\theta)$ is the Fisher Information, and $\widehat{I(\theta)}$ is the observed Fisher Information, i.e., $\widehat{I(\theta)} = -\frac{\partial^2 l_\Sigma(\theta)}{\partial \theta^2}|_{\hat{\theta},\hat{\Sigma}}$. The expression for $l_\Sigma(\theta)$ is given by (6). Some details on the asymptotic theory of non-linear model with a stationary linear process for the error can be found in Chapter 6 of Seber and Wild (2005). The modelling of the ARMA process will be of secondary importance with respect to the propulsive force. In the meantime, it is of primary importance to compare embedded models for the propulsive force. This can be done only by through an appropriate accounting for Σ.

Modelling for a particular speed profile. We will study in detail the fitting procedure for Subject 3, i.e., the blue velocity profile of Figure 1. The fitting strategy consists in choosing a complex model for the propulsive force while ignoring the dependence, and then identifying an appropriate ARMA structure from residuals. Finally, from this ARMA structure we will try to fit several embedded models and simplify, if possible. In Figure 8, there are four competitive models with $N_{\mathrm{p}} = 3$ fitted using the `nls()` command. The four models are defined below:

- Model 0 (Blue curve) corresponds to the more general form of (3) for $N_{\mathrm{p}} = 3$ and $|\theta| = 9$.

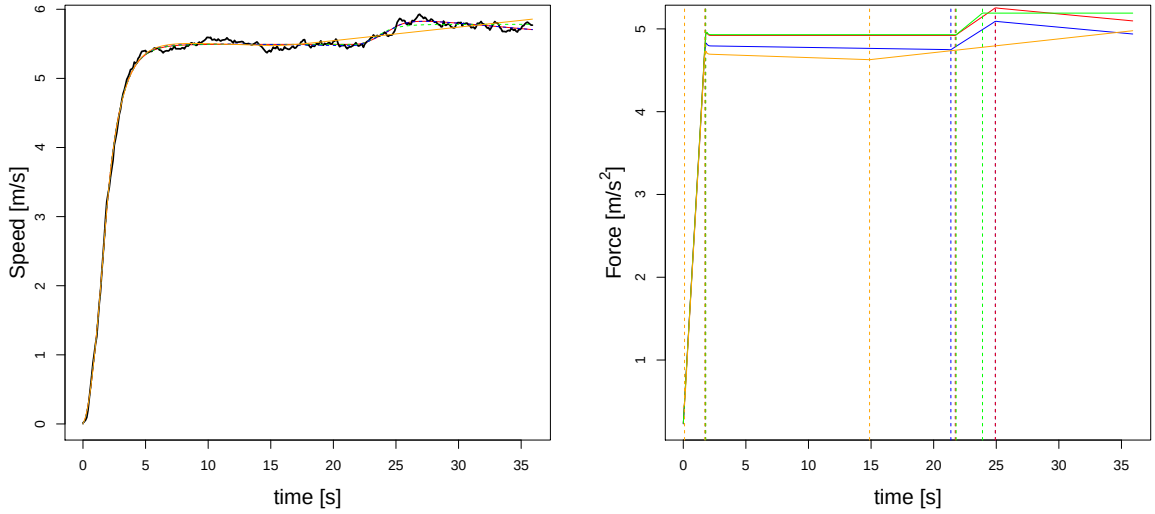

Figure 8: (Left) Fitted velocity profiled and the data (using an i.i.d. hypothesis on the residual). (Right) Propulsive force associated to fitted velocity profile. Dotted lines correspond to changing times $\hat{\eta}_k$. Subject 3.

- Model 1 (Red curve) constrains Model 0 by imposing $f_1 = f_2$, i.e. the propulsive force satisfies $f(t;\theta) = f_1$ for all $t \in [\eta_1, \eta_2)$.
- Model 2 (Green curve) constrains Model 0 by imposing $f_1 = f_2$ and $f_4 = f_5$, i.e. the propulsive force satisfies $f(t;\theta) = f_1$ for all $t \in [\eta_1, \eta_2)$ and $f(t;\theta) = f_4$ for all $t \in [\eta_4, T)$.
- Model 3 (Orange curve) constrains Model 0 by imposing $f_0 = f_1$, i.e. the propulsive force satisfies $f(t;\theta) = f_0$ for all $t \in [0, \eta_1)$.

These models answer qualitative questions concerning the profiled velocity. Indeed, Model 1 evaluates whether the speed should be constant on Figure between roughly 5 to 20 s 8 (Left). In addition, Model 2 evaluates whether the speed in the last phase between 25 to 35 s should also be constant. Moreover, Model 3 evaluates whether the propulsive force should be constant during initial acceleration. For Model 1 and 2 statistical testing is helpful. For Model 3, the closeness of the changing time $\hat{\eta}_1$ to 0 indicates that the propulsive force $f(t;\theta)$ should begin with an increasing phase (see the first orange dotted line on Figure 8). Imposing $f_0 = f_{\eta_1}$ restricts the numerical optimization procedure to pick η_1 so that this constraint has as little effect as possible. It strongly indicates that a constant propulsive force at the beginning of the race is not expected.

Model diagnostic From Model 0's residuals, we identified an ARMA(3,3) process. The ARMA model is fitted using `arima()` command. In Figure 9, the fitted error model shows adequate features to account for the correlation structure in the data. This error process is used to re-fit the whole model with the `optim()` command by exploiting the likelihood of (6). The final optimization of the whole model uses $\hat{\theta}$, as an initial guess, and $(\hat{\phi}, \hat{\kappa})$ found by numerical optimization of the NLS model and the ARMA model separately. From $\widehat{\epsilon_t} \sim \text{ARMA}(\hat{\phi}, \hat{\kappa})$, an estimate of the process covariance matrix $\widehat{\Sigma}$ is computed. The auto-covariance function of $\hat{\epsilon}_t$, denoted $\hat{\gamma}(h)$, $h = 0, 1, ..., N_{obs} - 1$, leads to an estimator

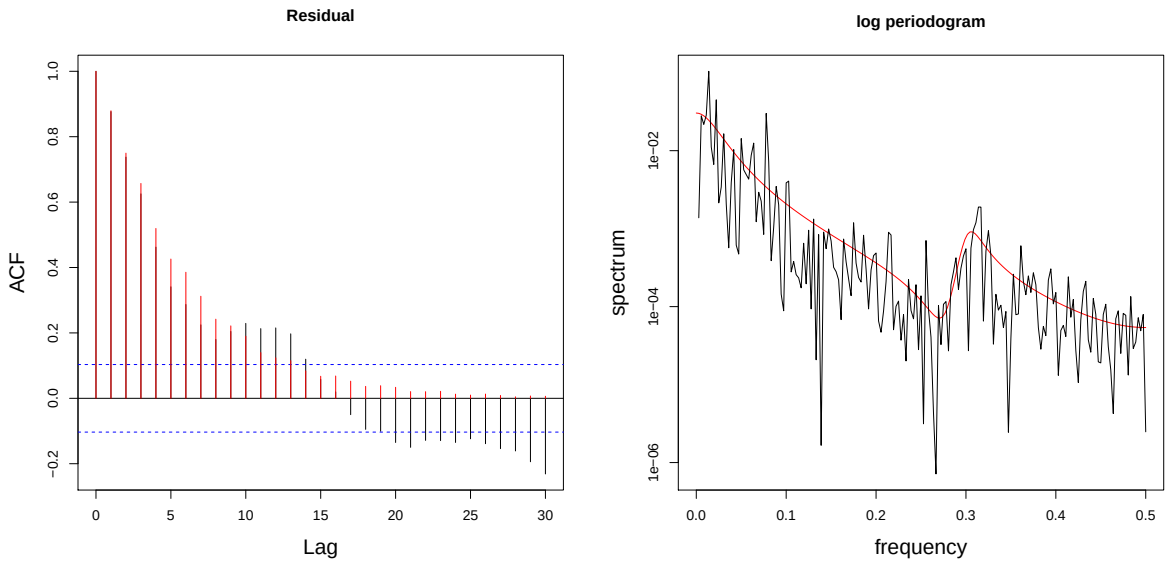

Figure 9: (Left) ACF for Model 0 s residuals in black and for fitted ARMA(3,3) in red. (Right) Periodogram for Model 0 s residuals in black and for fitted ARMA(3,3) in red. Subject 3.

of the covariance matrix defined by $\widehat{\Sigma}_{ij} = \hat{\gamma}(|i-j|)$ for $i, j = 0, 1, ..., N_{obs} - 1$. In Figure 10, a diagnostic plot for Model 1 with ARMA(3,3) is available. If the error model is appropriate, we should have $\epsilon_t \sim \mathcal{N}(0, \widehat{\Sigma})$. The innovation of Figure 10 corresponds to the transformed residuals $w_t = \widehat{\Sigma}^{-\frac{1}{2}} \epsilon_t$, and does not present a strong correlation structure, although the QQ-plot shows a good adequacy with the normality assumption. Similar diagnosis can be found using Model 0, 2 or 3.

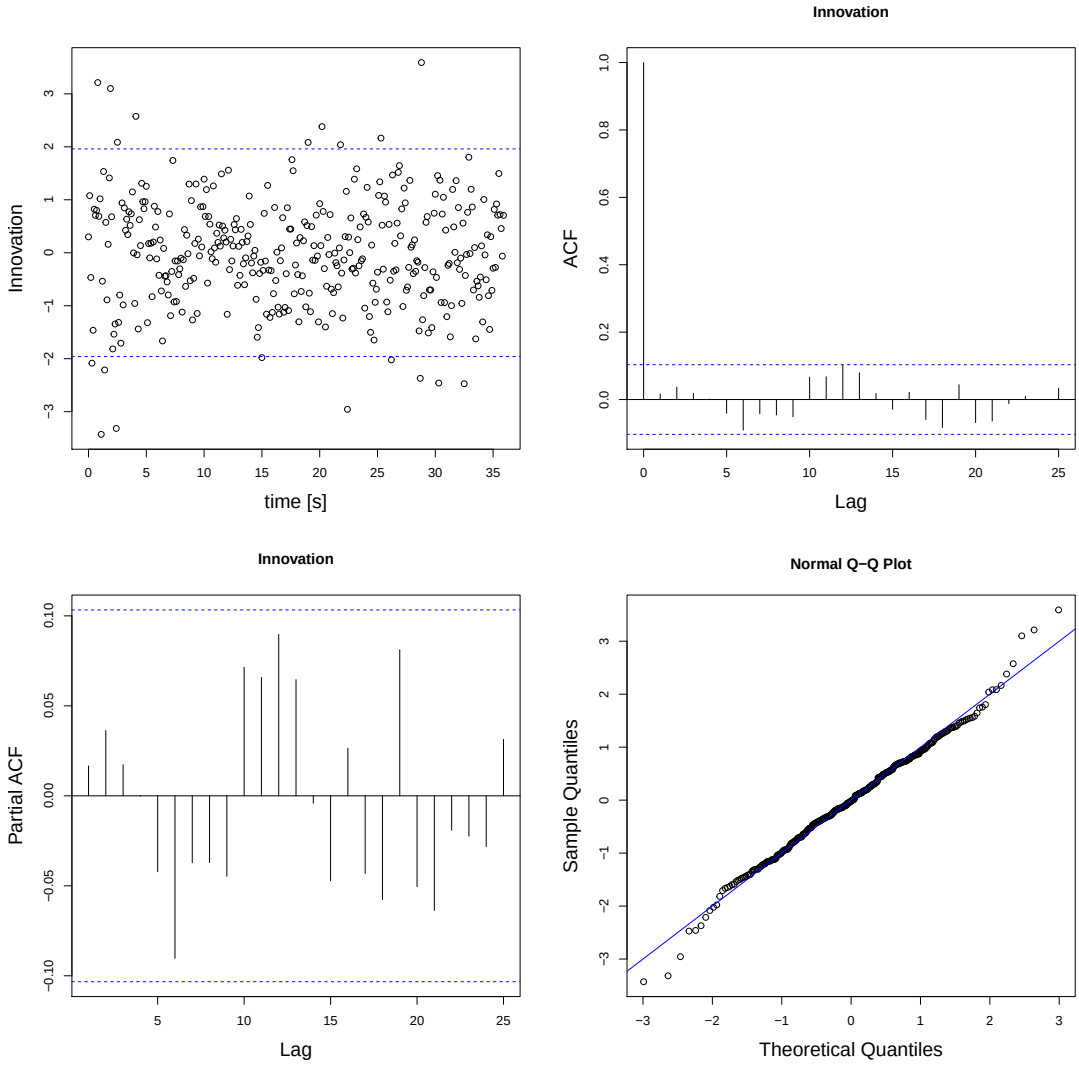

Figure 10: Diagnostics for correlation and normality for Model 1 with ARMA(3,3) error process.

	N_{para}	$\hat{l}$	$\hat{D}$	P-value	AIC	$\hat{l}^*$	$\hat{D}^*$	P-value*
Model 0	15	865.444	-	-	−1700.9	574.822	-	-
Model 1	14	864.734	1.420	0.233	−1701.5	570.051	9.544	0.002
Model 2	13	861.726	7.436	0.024	−1697.5	534.724	80.197	0.000
Model 3	14	854.443	22.003	0.000	−1680.9	405.879	337.888	0.000

Table 2: Table of likelihood statistics for comparison of models. Deviance statistics $\hat{D}$ are computed in comparison to Model 0. For columns marked by *, statistics are computed using an i.i.d. hypothesis. N_{para} is the total number of parameters for the mean model and the ARMA process. $\hat{l}$ is the log-likelihood.

Model selection On Table 2, the likelihood statistics and deviance for Model 1, 2 and 3 are presented for i.i.d hypothesis and ARMA(3,3) residuals. Table 2 highlights that neglecting the correlation in residuals leads to an excessively complicated model. For an ARMA modelling of residuals, Model 0 is not significantly better than Model 1 with respect to the likelihood ratio test. However, under an i.i.d. hypothesis Model 0 is significantly better than all simpler models. In addition, Model 2 becomes more competitive under ARMA modelling of residuals. Model 3 appears to be a poor choice in both cases. It confirms the interpretation given from Figure 8 (Right).

The interpretation of statistical tests in terms of the speed profile to Figure 8 (Left) appears to be coherent. Model 0 is not significantly better than Model 1 meaning that the first phase, $t \in (5, 20)s$, should be considered as constant speed, while the second phase, $t \in (25, 35)s$, should be considered as decreasing speed phase. For Model 1, the estimated values for $\theta = (\tau, f_0, f_{\eta_1}, f_{\eta_2}, f_{\eta_3}, \eta_1, \eta_2, \eta_3)$ and (ϕ, κ) can be found in Tables 3 and 4. AIC supports the choice of Model 1 as the preferred one.

	Estimate	Std Err
τ	1.158	0.036
f_0	0.117	0.114
f_{η_1}	4.745	0.144
f_{η_3}	5.063	0.159
f_T	4.920	0.157
η_1	1.681	0.047
η_2	21.853	0.309
η_3	24.944	0.459

Table 3: **Model 1**. ODE parameters' estimates and standard errors (computed using (17)). Subject 3.

	Estimate	Std.Err
ϕ_1	0.262	0.058
ϕ_2	-0.354	0.060
ϕ_3	0.697	0.053
κ_1	0.790	0.064
κ_2	0.968	0.055
κ_3	0.400	0.055

Table 4: **Model 1**. ARMA process's estimates and standard errors (computed using (17)). Subject 3.

Simultaneous Confidence Band for fitted speed profile The previous procedure provides estimates and uncertainty quantification for the physiological parameters τ, and the propulsive force $f(t;\theta)$. Based on these results, we can compute a simultaneous confidence interval for $\hat{v}(t) := v(t, \hat{\theta})$ based on Gsteiger et al. (2011). The difficulty is that $\hat{v}(t)$ is a non-linear function of parameters $\hat{\theta}$. To get around this problem, two levels of approximation are used in Gsteiger et al. (2011). First, a linearization of $\hat{v}(t)$, and then the normal approximation of the estimator's distribution $\hat{\theta} \sim \mathcal{N}(\theta, \widehat{I(\hat{\theta})}^{-1})$:

$$\begin{aligned} v(t,\hat{\theta}) - v(t,\theta) &\approx \nabla_\theta v(t;\hat{\theta})^T(\hat{\theta} - \theta) \\ &\overset{d}{\sim} \mathcal{N}(0, \nabla_\theta \hat{v}(t)^T \widehat{I(\hat{\theta})}^{-1} \nabla_\theta \hat{v}(t)). \end{aligned}$$

where $v(t) := v(t;\theta)$ is the true velocity profile. The previous computation gives an approximate distribution of $\hat{v}(t)$. The resulting distribution is time-dependent in the mean and in the variance. In order to obtain the confidence band for $\hat{v}(t)$ and $t \in [0, T]$, one has to find the critical value d_α so that:

$$P\left(v(t) \in \hat{v}(t) \pm d_\alpha \sqrt{\nabla_\theta \hat{v}(t)^T \Sigma_\theta \nabla_\theta \hat{v}(t)} \text{ for all } t \in [0,T]\right) = 1 - \alpha$$

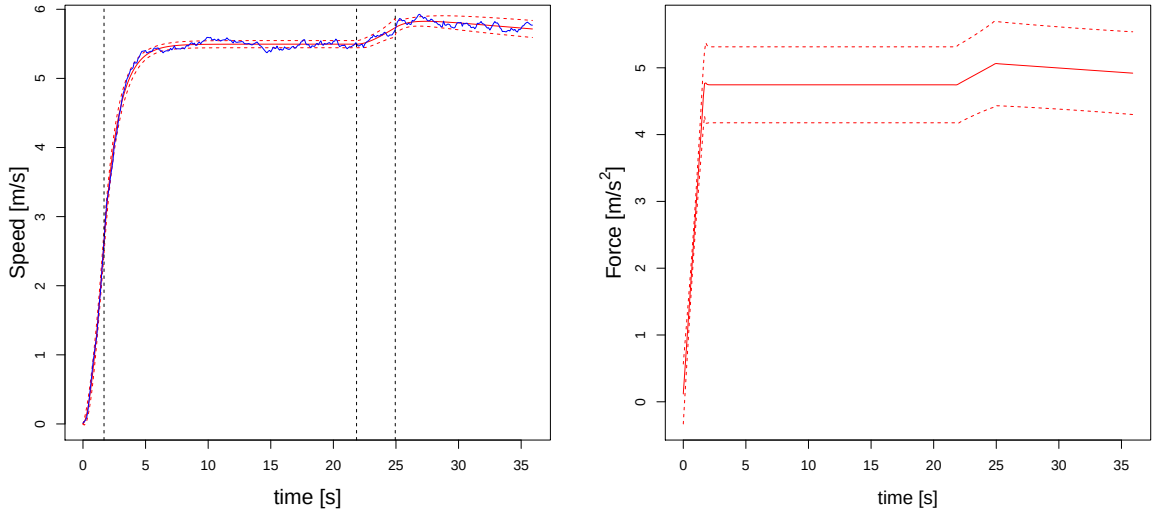

Figure 11: Simultaneous 95% confidence band for the speed profile (Left) and the propulsive force (Right). The dashed lines correspond to changing times $\hat{\eta}_1$,$\hat{\eta}_2$ and $\hat{\eta}_3$

where $\Sigma_\theta := \widehat{I(\hat{\theta})}^{-1}$. The critical value is obtained with the Cauchy-Schwarz inequality on the re-scaled fitted velocity profile $\hat{v}(t)$ (see equation 3 in Gsteiger et al. (2011)); and

$$d_\alpha = \sqrt{q_{\chi^2_p}(1-\alpha)}$$

where $q_{\chi^2_p}(1-\alpha)$ is the $1-\alpha$ quantile of a chi-squared distribution with $p = |\theta|$. Although the degrees of freedom p does not depend on the complexity of ARMA residuals, the auto-correlation is taken into account in Σ_θ because the Fisher Information is derived from the Hessian of $\hat{l} = l(\hat{\theta}, \widehat{\Sigma})$. The fitted speed profile's gradient $\nabla_\theta \hat{v}$ is computed with finite-difference method using `fdHess` command from `nlme` package.

The confidence band is computed for Model 1 and the result is displayed in Figure 11. A similar confidence band is derived for the fitted propulsive force $\hat{f}(t) := f(t;\hat{\theta})$. Since $\nabla_\theta \hat{f}(t)$ is computed with finite-difference approximation, the expression of (3) for the propulsive force is not appropriate for numerical purposes. Instead of the indicator function $\mathbb{1}[\eta_{k-1}, \eta_k]$, we use a smoothed version of this function:

$$\tilde{\mathbb{1}}[\eta_{k-1}, \eta_k](t) = \frac{1}{1+\exp(-\gamma(t-\eta_{k-1}))} - \frac{1}{1+\exp(-\gamma(t-\eta_k))} \tag{18}$$

where γ is a fixed constant that controls the closeness of the smoothed approximation $\tilde{\mathbb{1}}[\eta_{k-1}, \eta_k]$ of the indicator function $\mathbb{1}[\eta_{k-1}, \eta_k]$. The smooth version and the indicator function are illustrated in Figure 12 (Left). This approximation is slightly abusive since we fitted the ODE's explicit solution (given by (15)) for a propulsive force with indicator functions. However, it should not have an important incidence on the simultaneous band, since approximating the step function, $\nabla_\theta \hat{f}(t)$, by a smoother version has no influence outside a small region around the discontinuities. Figure 11 shows some important features. First, the uncertainty on the propulsive force $\hat{f}(t)$ is much larger than on $\hat{v}(t)$. This seems rational because the parametric form for $f(t;\theta)$ is unknown, and the confidence band accounts for it appropriately. Besides, this large uncertainty on $\hat{f}(t)$ does not reverberate

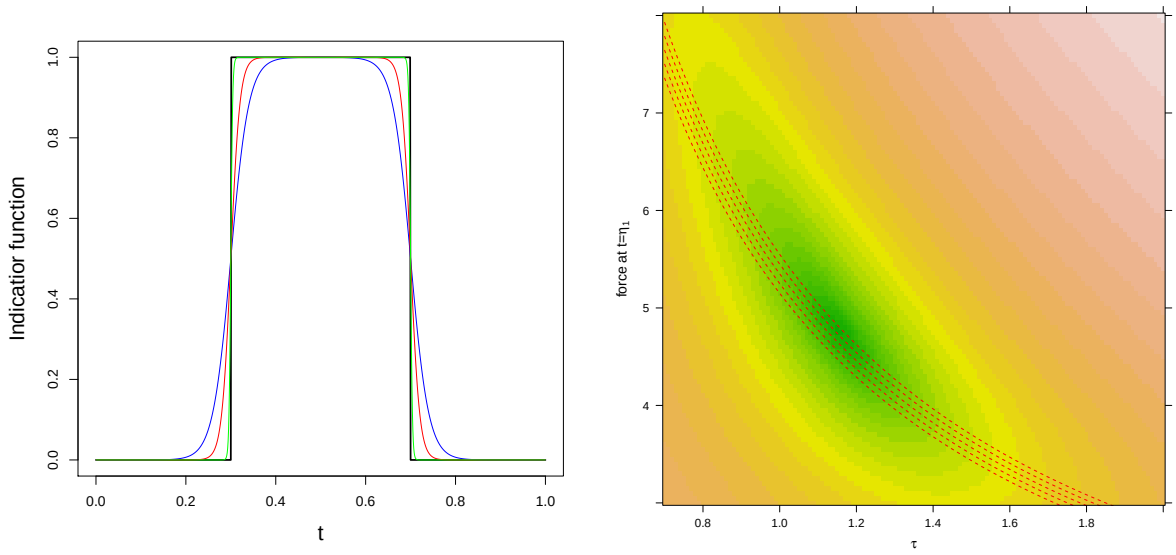

Figure 12: (Left) Approximation of the indicator function with a sum of two sigmoid functions (see (18)) for $\gamma = 50, 100, 500$. (Right) Likelihood levelplot for Model 1 with ARMA residual, in function of τ and f_{η_1}. The dotted red lines correspond to the curves $f_{\eta_1} = C/\tau$, where $C \approx 5.45\ m.s^{-1}$, i.e., the mean value of velocity profile for $t \in (5, 20)s$.

on $\hat{v}(t)$. This is due to an ambiguity on the value of $f(t;\theta)$ during a constant phase of speed: $\dot{v} = 0$. Along a constant phase one notes that $v(t) = Cst = f\tau$ implies an ambiguity on the value of f. In Figure 12 (Right) we observe that the optimization surface around the minimum is more stretched along the force direction. This ambiguity on the force during a constant phase does not impact speed uncertainty thanks to the strong correlation between $\hat{\tau}$ and $\hat{f}$. If t is such that $\hat{v}(t) = \hat{\tau}\hat{f} = Cst$, then the approximate variance of the fitted value in such regions is:

$$\begin{aligned}\mathrm{Var}(\hat{v}(t)) &= \nabla_\theta \hat{v}(t)^T \Sigma_\theta \nabla_\theta \hat{v}(t) \\ &= \partial_\tau \hat{v}\sigma_f + \partial_f \hat{v}\sigma_\tau + 2\partial_f \hat{v}\partial_\tau \hat{v}\sigma_{\tau f} \\ &= \hat{f}\sigma_\tau + \hat{\tau}\sigma_f + 2\sigma_{\tau f}\hat{f}\hat{\tau},\end{aligned}$$

where σ_τ, σ_f and $\sigma_{f\tau}$ are the variance and covariance for $\hat{\tau}$ and $\hat{f}$. The covariance between $\hat{\tau}$ and $\hat{f}$, $\sigma_{\tau f}$, is strongly negative and will contribute in reducing $\mathrm{Var}\{\hat{v}(t)\}$. For instance, for Model 1 the correlations between f_{η_1}, f_{η_3} and f_{η_T} are of order -0.98. This strong negative correlation between $\hat{\tau}$ and $\hat{f}$ appears clearly on Figure 12 (Right), and is associated to the constant phase for $t \in (5, 20)s$.

5.2 GP method

This section explores GP method's properties for different values of λ. We illustrate two cases for which the λ must be chosen differently. The first exploits the ODE Model 0 from the previous section. It is a well-specified model in the sense that Model 0's propulsive force leads to an accurate description of the data. In this case, we will try to confirm that, for larger value of λ, the GP method performs as accurately as the NLS method. Besides, we investigate the possibility to use $\hat{\theta}(\lambda)$ for low value of λ as a clever initial guess for larger value of λ. The second case investigates a simpler model than Model 0 but under a

relaxation of the ODE model, i.e., for low to mid values of λ. The motivation remains to simplify propulsive force modelling while keeping an appropriate fit to the data. We will investigate uncertainty behavior for this mis-specified model, as well as the interpretation of ODE parameter, in this relaxed framework.

Explicit formulas for the generalized smoothing matrix In Section 3.1, we explained the procedure used to obtain a relaxed solution of an ODE. While the routines of the `CollocInfer` package are exploited in Section 3.3, in this section we will use our own fitting procedure, because the relaxed solution problem defined by the minimization of (8) has an explicit solution, whose derivation can be found in Appendix A.1. The basis functions of the cubic spline space with knots at each sampling time, S_K^4, are denoted $\{\phi_i\}_{i=1}^{N_{basis}}$. From B-spline space properties, the number of basis functions is $N_{basis} = N_{obs} + 2$. For given values of θ and λ, the optimal coefficient $c \in \mathbb{R}^{N_{basis}}$ solution of (8) is:

$$\hat{c}(\theta, \lambda) = (\Phi^T \Phi + \lambda \Omega_\theta)^{-1} (\Phi^T y + \lambda b_\theta),$$

where Ω_θ is a $N_{basis} \times N_{basis}$ positive definite matrix[3], $\Phi_{ij} = \phi_j(t_i)$ with $i = 1, ..., N_{obs}$, $j = 1, ..., N_{basis}$ and $b_\theta \in \mathbb{R}^{N_{basis}}$. It turns out that Ω_θ depends only on τ, while b_θ depends on τ and on $f_0, ..., f_T, \eta_1, ..., \eta_{N_p - 1}$. The fitted values $\hat{v}(\theta, \lambda) = \Phi \hat{c}(\theta, \lambda)$ could be re-written:

$$\hat{v}(\theta, \lambda) = H_{\lambda,\theta} y + k_{\theta,\lambda} \tag{19}$$

where $H_{\lambda,\theta} = \Phi(\Phi^T \Phi + \lambda \Omega_\theta)^{-1} \Phi^T$ is a smoothing matrix, and $k_{\theta,\lambda} = \Phi(\Phi^T \Phi + \lambda \Omega_\theta)^{-1} \lambda b_\theta$ is a N_{obs} vector. The vector of fitted relaxed solutions $\hat{v}(\theta, \lambda)$ is the sum of filtered data $H_{\lambda,\theta} y$ and of a shift $k_{\theta,\lambda}$ that depends on the propulsive force. While in the spline smoothing context the matrix $H_{\theta,\lambda}$ is familiar, the shift vector is unusual. This term comes from the definition of the roughness penalty as a non-homogeneous problem $L_\theta v = f_\theta$, as opposed to spline smoothing, where the roughness penalty is defined from a homogeneous problem $D^2 v = 0$. The smoothing matrix $H_{\lambda,\theta}$ has similar properties as in the spline smoothing context: the equivalent degrees of freedom vary continuously with λ between 1 and N_{obs}. For spline smoothing, the degrees of freedom vary between 2 and N_{obs}. Note that this difference comes from the kernel's dimension of linear operator defining the homogeneous ODE associated to the roughness penalty. In the spline smoothing context operator's kernel is $\mathrm{Ker}(D^2) = \mathrm{Span}\{1, t\}$, while for our ODE $\mathrm{Ker}(L_\theta) = \mathrm{Span}\{\exp(-t/\tau)\}$. In Figure 13 (Left), we observe the degrees of freedom of $H_{\lambda,\theta}$ decaying to 1 as $\lambda \to +\infty$. For a large λ, the interpretation is that the data y is"projected" into a subspace of almost dimension 1. From Figure 13 (Right), we observe that the eigenvector associated to the largest eigenvalue satisfies $v_1 = \alpha \exp(-t/\tau)|_{\{t_i\}_{i=1}^{N_{obs}}}$. The eigenvectors v_1,v_2 and v_3 are displayed for $\lambda = 10000$, but are essentially the same for all λ. Only the eigenvalues μ_i change for different values of λ. Besides, for $\lambda \geq 5000$ we have $\frac{\mu_1}{\sum_{i=1}^{N_{obs}} \mu_i} > 95\%$, therefore the subspace of dimension 1 onto which the data are "projected" coincides approximately with the eigenspace spanned by v_1. For a large λ, the following approximation becomes reasonable:

$$H_{\lambda,\theta} \approx \mu_1 v_1 v_1^T$$

and this approximation shows that data's smoothing essentially filters out signal-components that are not a linear combination of $v_1 = \exp(-t/\tau)|_{\{t_i\}_{i=1}^{N_{obs}}}$: $H_{\lambda,\theta} y \propto v_1$. In the following, the formula of (19) is used to apply the algorithm procedure described in Section 3.2.

[3]The exact definition is given in the Appendix.

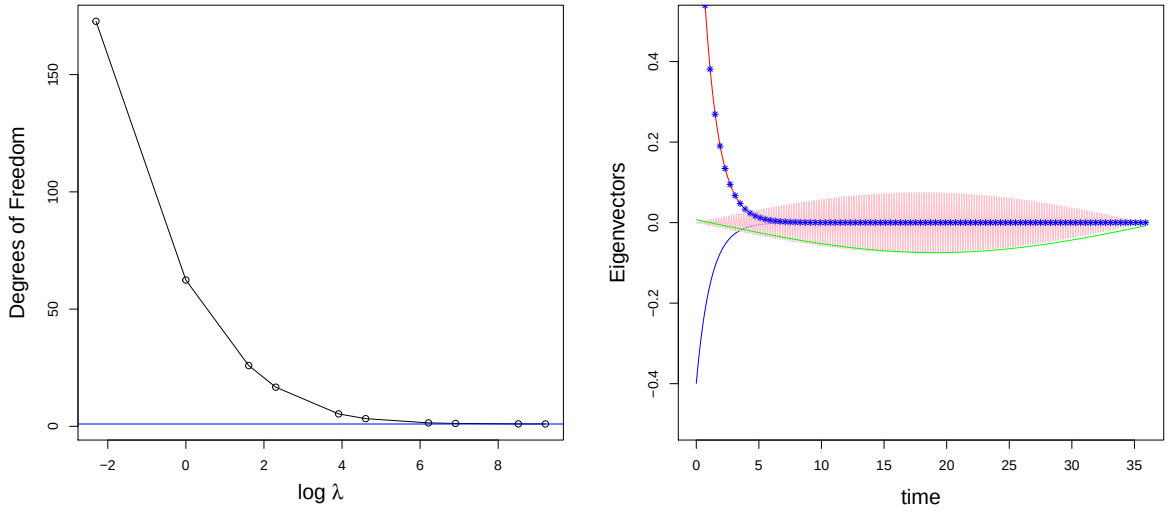

Figure 13: (Left) Degrees of freedom defined as the trace of $H_{\lambda,\theta}$ for $\lambda = 0.1, ..., 10000$. (Right) Eigenvectors of $H_{\lambda,\theta}$ associated to the three largest eigenvalues $\mu_1 > \mu_2 > \mu_3$, with v_1 in Blue, v_2 in Pink and v_3 in Green. The red curve is $\exp(-t/\tau)$. We used the same data as in NLS method and $\tau = 1.15$. The blue stars correspond to the vector $-2.48v_1$.

More specifically, at step 1 the formula is used to compute a relaxed solution for each updated value of θ. The minimization of $H(\theta)$ can be done with the `nls()` command, by defining a function that returns a relaxed solution defined by (19). To account for the auto-correlation in the residuals, we identify an appropriate ARMA model and its correlation structure $\hat{\Gamma}$, and apply the NLS algorithm for the re-scaled data $\hat{\Gamma}^{-\frac{1}{2}}y$ and the re-scaled relaxed-solution $\hat{\Gamma}^{-\frac{1}{2}}\hat{v}(\theta, \lambda)$. This method is less efficient than in Section 5.1, as the ARMA structure and θ are not estimated simultaneously. However, the optimization of $H(\theta)$ is already slow because each relaxed-solution computation requires $O(N_{obs}^2)$ operations. For optimization simplicity and because of lack of time, we prefer to apply the `nls()` command on the re-scaled data and model.

Model 0 with GP method The theory developed by Qui and Zhao (2010) suggests that the GP method should perform as well as the NLS method in terms of asymptotic behavior for θ. In addition, the GP method has the advantage of potentially avoiding local minimas, by performing the fitting process for a gradually increasing sequence of values of λ. The first statement is proven in Qui and Zhao (2010), while the second one is conjectured in Ramsay et al. (2007) and Qui and Zhao (2010). We adjusted the Model 0 of Section 5.1 to the data for increasing values of λ. For a fixed value of λ, the estimated standard deviation is computed from the curvature of the data-fitting criterion $H(\theta; \lambda)$ as explained by (11). The exact definition of $H(\theta; \lambda)$ can be found in (10). To define the likelihood g_Σ of (10), the correlation structure Σ of the ARMA(3,3) process defined by the coefficients on Table 4 was used. The results are displayed for $\hat{\tau}$ and $\hat{f}_{\eta_1}$ in Figure 14 (Top-Left and Top-Right). Similar patterns are observed for other parameters. The standard deviation is larger for smaller values of λ, and for $\lambda > 1000$ it remains unchanged. This pattern is typical of the GP method, and has already been discussed in Section 3.3. For large enough λ, the collocation approximation $v(\theta, \lambda) = \Phi c(\theta, \lambda)$ stops

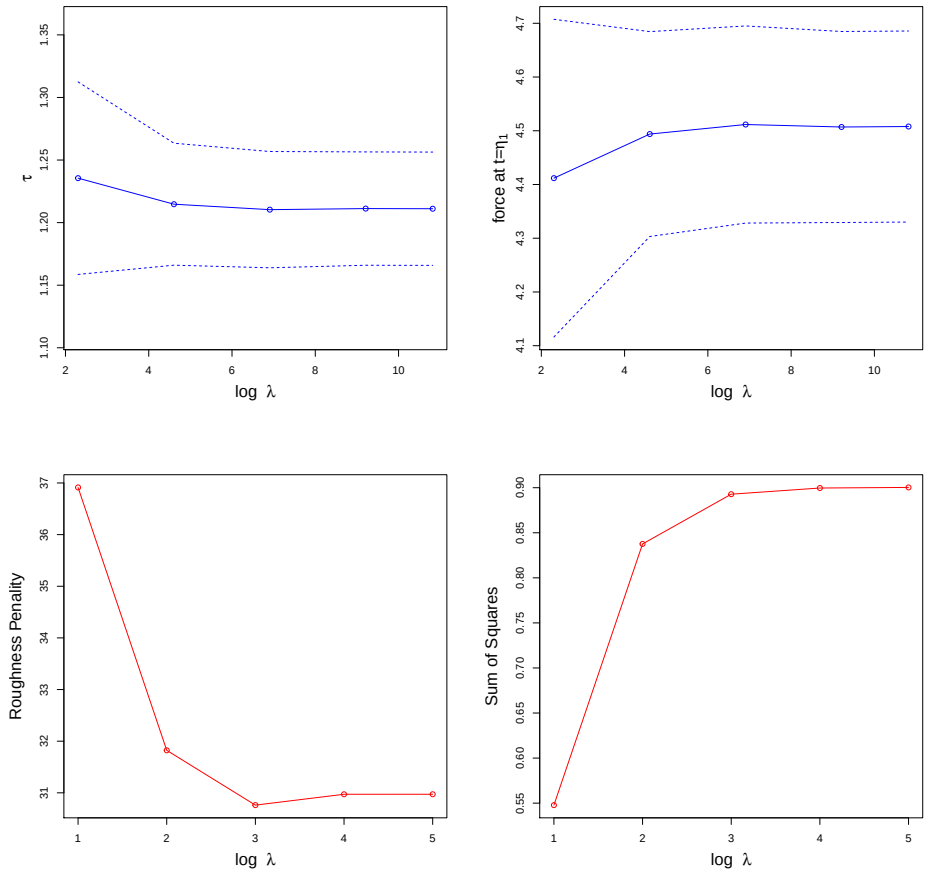

Figure 14: (Top-Left and Top-Right) The dotted lines are the estimated values of τ and f_{η_1} for $\lambda = 10, ..., 50000$. The dashed lines correspond to estimated standard deviation using observed Fisher Information. (Bottom-Left and Bottom-Right) Roughness penality and residuals sum of squares as functions of λ.

improving and essentially reaches a good approximation of ODE's analytic solutions. The GP method behaves like the NLS method, as expected. Indeed, the estimate $\hat{\theta}$ is such that $v(\hat{\theta}, \lambda)$ is the closest fit to the data, while $v(\hat{\theta}, \lambda)$ is almost like the exact solutions used for NLS method. Note that for $\lambda > 1000$, the estimated standard deviation are $\widehat{SD}(\hat{\tau}(\lambda)) = 0.17$ and $\widehat{SD}(\hat{\tau}(\lambda)) = 0.044$. This is very similar to the result obtained in Table 3. The estimated values are a little different but it comes from the choice of Model 0 instead of Model 1 for Table 3.

In Figure 14 (Bottom-Left and Bottom-Right), the roughness penalty $\text{PEN}(\theta, \lambda) = \int_0^T \{L_\theta v(s) - f(s;\theta)\}^2 \, ds$ and the residual sum of squares are displayed for $\theta = \hat{\theta}(\lambda)$. The observed curves summarize the previous discussion. For $\lambda > 1000$, the roughness penalty flats-out as the collocation approximation saturates, while the sum of squares rises. The sum of squares varies only between 0.55 and 0.90 with λ, indicating that the ODE model is appropriate for the data. According to Qui and Zhao (2010), in such a context, the choice of λ must be guided by the quality of the estimation. If the estimated value and its confidence interval stop changing for large λ then we may stop increasing it. From (19), one observes that the dependence on θ is almost cancelled when $\lambda \to 0$. This implies the flatness with respect to θ of the optimization surface $H(\theta; \lambda)$ for small values of λ. In Figure 15, the latter discussion is confirmed since the fitting-surface varies much less

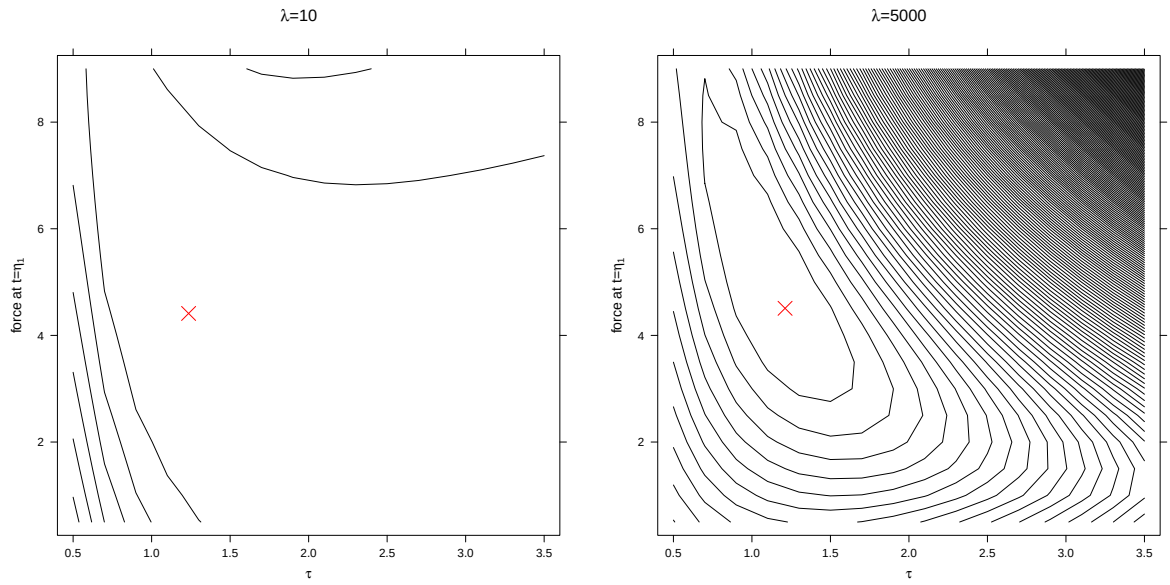

Figure 15: Contour plot of the optimization surface $H(\theta;\lambda)$. The contours are defined for the same levels going from 0 to 1.4e5 by 700. The red-cross corresponds to estimated values.

around the estimated value for $\lambda = 10$ than for $\lambda = 5000$. This may present an advantage for two reasons. First, it tends to erase local-optima while the global minimum for a large value of λ will not be far. Hence, the GP method fitted with an increasing sequence of λ could remove the difficulty of finding an appropriate initial guess. Secondly, the fitting-surface could account for a larger uncertainty on $\hat{\theta}(\lambda)$ for lower values of λ, recalling that the estimated standard deviation of (11) is defined with the curvature of $H(\theta;\lambda)$ around the minimum.

Simpler Model with GP In this paragraph, we are going to use a simpler model and compare its performance with the NLS method for Model 0. The model used presents only one changing time $N_{\mathrm{p}} = 1$. In Figure 8 (Right), as a consequence, the local increase of the propulsive force $f(t;\theta)$, between $22s$ and $27s$, is erased. The parameter becomes $\theta = (\tau, f_0, f_{\eta_1}, f_T, \eta_1)$. As the model is simpler, the error process gets more complicated. Instead of an ARMA(3,3) we need an ARMA(4,4) to remove the correlations in the residuals. The results for the estimated values are displayed in Figure 16. In Figure 16 (Bottom-right), for $\log(\lambda) \leq 3$ the sum of squares and penalty scores inform us that the fitted curve adjusts better to the data while the ODE model is not strictly trusted. In the meantime, the estimators standard deviation does not increase significantly before $\log\lambda < 0$. The pattern is slightly different than in Figure 14 (Top). Before the increase of the standard deviation, the estimates change their values. This change indicates that the ODE model is not appropriate for the data, but it is informative since the standard errors remain constant during this change. It is interesting to compare the sum of squares/penalty pattern with the estimation and uncertainty pattern for the simple and complex model. For the complex model in Figure 14 (Top), the uncertainty increase appears instantaneously when the roughness penalty starts to rise. On the other hand for the simple model, this uncertainty increase arises after the roughness penalty has exploded and the sum of squares is close to 0.

In Figure 17 (Left), the velocity fit for the simple model and the associated estimated

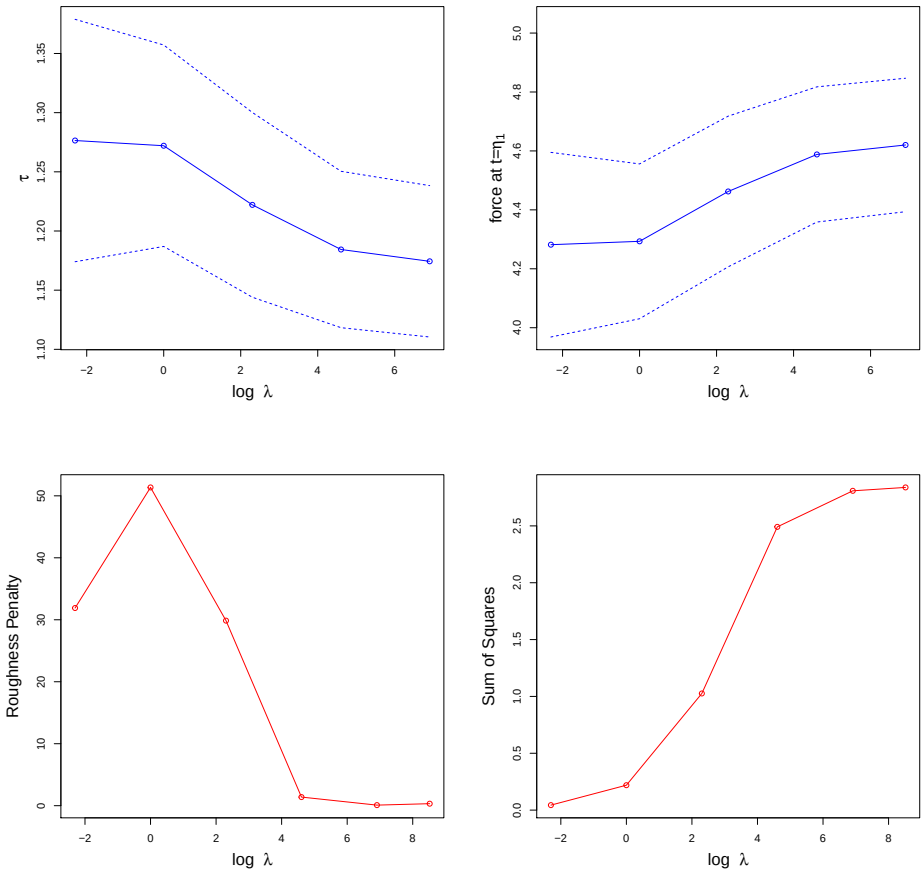

Figure 16: (Top-Left and Top-Right) Estimated values with GP method and respective estimated standard errors. (Bottom-Left and Bottom-Right) Roughness penality and residuals sum of squares as functions of λ.

propulsive force for different λ are displayed. These estimated velocity and force profiles are compared to the Model 0 adjusted with the NLS method. The result is particularly interesting for $\log \lambda = 2.3$ (green curve). The fitted velocity profile $v(t; \hat{\theta}, \lambda)$ is very close to Model 0. In addition, the estimated propulsive force compromises the one found for the Model 0. It is an example of a mis-specified model for the propulsive force $f(t; \theta)$ that performs well in terms of data-fitting (as Model 0 does), and in addition estimates $f(t; \theta)$ in a coherent way with respect to the well specified model.

From the two previous paragraphs, we may first conclude that the approximation $\mathrm{Cov}(\hat{\theta}(\lambda)) \approx \frac{d^2 H(\theta;\lambda)}{d\theta^2}^{-1}|_{\theta=\hat{\theta}}$ gives an appropriate result for a large λ in the sense that standard errors will correspond to the NLS ones for the same model. This result is predicted by the asymptotic theory in Qui and Zhao (2010). Secondly, for intermediate and low values of λ, this covariance approximation gets larger when the ODE model becomes uninformative with regard to the fitting process. Hence, it does not strictly account for a departure from the ODE model. More specifically, for small values of λ (i.e. a low value for the sum of squares and a high value for the penalty), the standard errors do not rise if the ODE model remains informative. Thirdly, the GP method for a simpler model, and at a value of λ balancing the sum of squares and roughness penalty, performs almost as accurately as in the case of a well-specified model. This

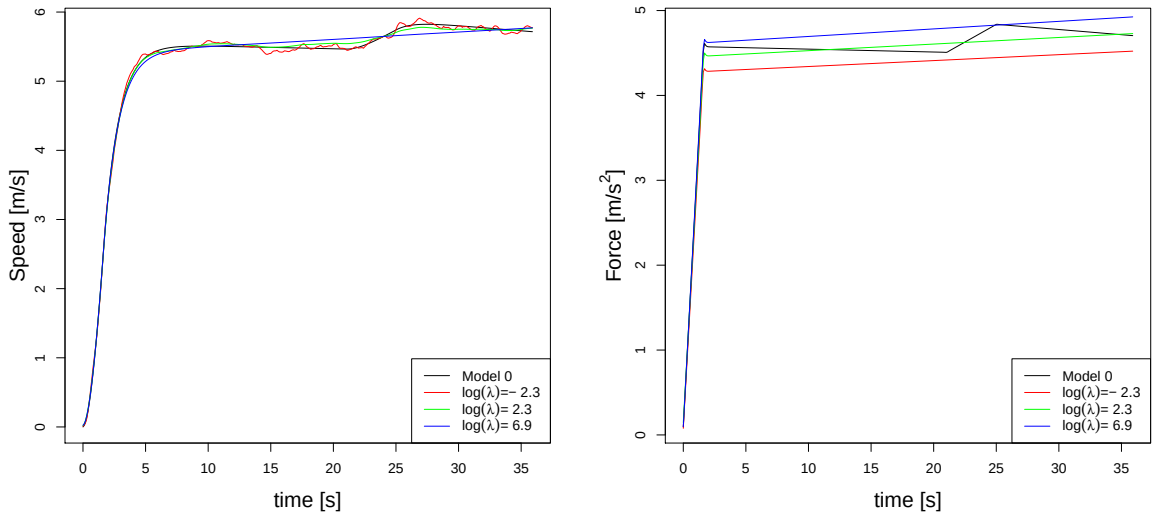

Figure 17: Simple model. (Left) The fitted velocity for the GP method with different value for λ. (Right) The estimated propulsive force with different values for λ.

latter result is interesting for modelling simplification. However, the asymptotic theory is not developed for intermediate values of λ. Therefore, there is no guarantee that the covariance estimate based on the curvature of $H(\theta;\lambda)$ is appropriate for an asymptotic normality approximation.

5.3 Two-step method

This section aims at showing that the two-step method may provide comparable estimates for θ as the ones provided in the previous sections. Due to a lack of time the two-step method has not been fully investigated. In particular, we will consider that the changing times η_k are given, instead of being estimated from data. We used Model 0 from the NLS section to compare the two-step method's performance. The main advantage of this method is the simplicity of the model-fitting, once non-parametric estimators $\hat{v}(t)$ and $\hat{\dot{v}}(t)$ are given. Model-fitting can be performed with the `lm()` command as described in Section 4.1. The design matrix X of this linear model becomes more complex than in Section 4.2.1, because there is a need to account for the different pieces N_{p} of the propulsive force. Indeed, the matrix X must account for groups corresponding to pieces of the propulsive force. The mesh on which the integral approximation (14) is performed will be noted $P = \{t_0 = 0 < t_1 ... < t_{N_{mesh}} = T\}$. Without going into details on the structure of X, it is clearer to state the R-formula used:

$$\texttt{Acceleration} \sim \texttt{Speed} + \sum_{k=1}^{N_{\mathrm{p}}} \texttt{I}_k + \texttt{Time*I}_k - 1$$

where each variable is an evaluation of the following functions: $\hat{\dot{v}}(t)$, $\hat{v}(t)$, $t\mathbb{1}[\eta_{k-1},\eta_k](t)$, and $\mathbb{1}[\eta_{k-1},\eta_k](t)$ on the mesh P, for $k = 1, ..., N_{\mathrm{p}}$. Using this R-formula with appropriate weights, the integral criterion $R^2_w(\theta)$ is approximated as described in (14). The weight function in $R^2_w(\theta)$ is $w(t) = 1$, and we use a trapezoidal rule for the quadrature formula. The non-parametric estimators $\hat{v}(t)$ and $\hat{\dot{v}}(t)$ are computed with smoothing spline for

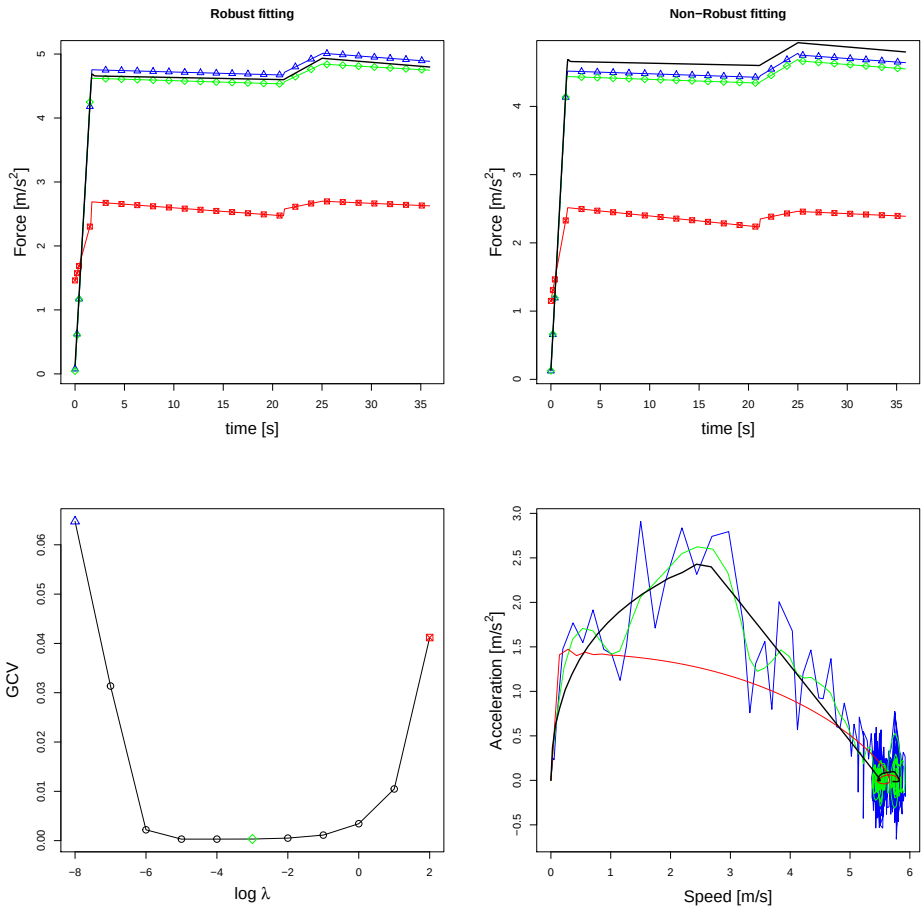

Figure 18: Blue, green, and red curves are related to the non-parametric estimators $\hat{v}$ and $\hat{\dot{v}}$ for values of λ given in the Generalized Cross Validation plot (Bottom-Left). Black curve corresponds to the Model 0 fitted with NLS. (Top) Estimated propulsive force for the two-step method, using or not a robust linear model fitting. (Bottom-Right) Phase-plane plot of non-parametric estimators and Model 0. Subject 3.

various values of the regularization parameter λ. The choice of λ may be guided through the Generalized Cross-Validation (GCV) criterion, but the results in Figure 18 (Bottom-Left) indicate that there is no clear choice for λ. The two-step method is performed for three values of λ highlighted on the GCV plot. In Figure 18 (Top), we observe that the estimated propulsive force is coherent with the NLS method for Model 0 provided that the data are not over-smoothed by the non-parametric estimators. An over-smoothed estimator of acceleration and speed distorts the initial acceleration phase in Figure 18 (Bottom-Right) (red-curve). Besides, the robust linear fitting (fitted with the `rlm()` command and the Huber ψ function) reinforces the coherency already mentioned. In Figure 18 (Bottom-Right), the oscillations observed are due to the non-parametric estimators in the first step. Their wiggling behaviors introduce a bias on the piecewise regression. The robust fitting tends to attenuate the effects of these undesirable oscillations. The improvement of the two-step method with respect to appropriate weighting of criterion $R^2_w(\theta)$ was already predicted from simulations in Section 4.2.2. The robust linear fitting chooses these weights depending on the data.

This example shows that the two-step method provides similar results as the NLS method for identical changing times. However, the estimation of these changing times

in the two-step method framework is challenging. From the phase-plane of Figure 18 (Bottom-Right), it appears difficult to distinguish the propulsive force's pieces. This problem must be further investigated, but it is already apparent that the modelling of the propulsive force in function of changing times η_k is not appropriate. Concerning the uncertainty quantification of the two-step method, a bootstrap strategy might be appropriate. In order to account for the error in each step, one should simulate the error process from the residuals $\epsilon = y - \hat{v}(t)|_{\{t_i\}_{i=1}^{N_{obs}}}$. For each simulation, the two-step procedure must be repeated entirely. Yet, this is not an issue since the two-step method is very fast, and all simulations can be performed with one vectorize computation as explained in Section 4.2.1.

5.4 Results for other runners

This section briefly presents the results obtained for other runners than Subject 3. The results of the NLS method, being the most relevant to our study, will be the only ones presented. The comparison procedure detailed in Section 5.1 seems to apply quite well for other runners. The only difference comes from the structure of the ARMA model used. For Subject 1, the simultaneous confidence interval for the fitted speed profile $\hat{v}(t)$ and the fitted propulsive force $\hat{f}(t)$ are presented in Figure 19 (Top). In addition, Figure 19 (Top) shows the residual's periodogram and the innovation ACF. The correlation model is ARMA(6,3), and its high auto-regressive order is needed to capture the quasi-periodic behavior observed in the residual. This behavior is not observed for Subject 3 in Figure 9 (Right), and it is not caused by the fitting process, since the peaks exist in the original data. A similar behavior is obtained for Subject 2. The fundamental frequency of the phenomenon $\omega = 1.17$ Hz strongly suggests a relation with the runner's foot pace. It must be recalled that LMAM's GPS equipment is installed on a cap over the runner's head. The runner's head slight vertical movements at each foot pace could induce measurement error. Well-trained runners tend to keep their head at the same height along the race. The fact that this periodicity is absent or very weak in previous modelling could make sense since Subject 3 is the best runner among the three. The relevant estimated values are given in Table 5. Note also that Subject 1 has a larger inertial drag $\hat{\tau} = 1.44\ s$ than Subject 3 $\hat{\tau} = 1.15\ s$, and this is consistent with the curves in Figure 1. Indeed, the inertial drag impacts the typical time of variation to reach a certain speed, and the blue curve goes faster to its cruise speed than the green one. From the analysis of Subject

	τ	f_0	f_{η_1}	f_{η_2}	f_{η_3}	f_{η_4}	f_T	η_1	η_2	η_3	η_4
$\hat{\theta}$	1.445	0.155	3.141	3.512	3.312	3.723	3.328	2.229	10.450	14.091	25.940
$\widehat{SD}(\hat{\theta})$	0.087	0.070	0.146	0.218	0.199	0.222	0.204	0.064	0.187	0.227	0.349

Table 5: ODE parameters' estimates and standard errors (computed using (17)). Subject 1.

1, we observe similar features as with Subject 3 except for the quasi-periodicity in the residual. Since the ARMA(6,3) model correctly captures the main periodicity associated to the foot pace, the simplification procedure of the propulsive force $f(t;\theta)$ could be attempted. However, for Subject 2 the situation is more complicated. In Figure 20 again, the residuals' periodogram shows a strong periodicity even though the first harmonic is stronger than for Subject 1. To capture both peaks, we need a higher order ARMA model as illustrated in Figure 20 for an ARMA(8,6). The simultaneous numerical optimization

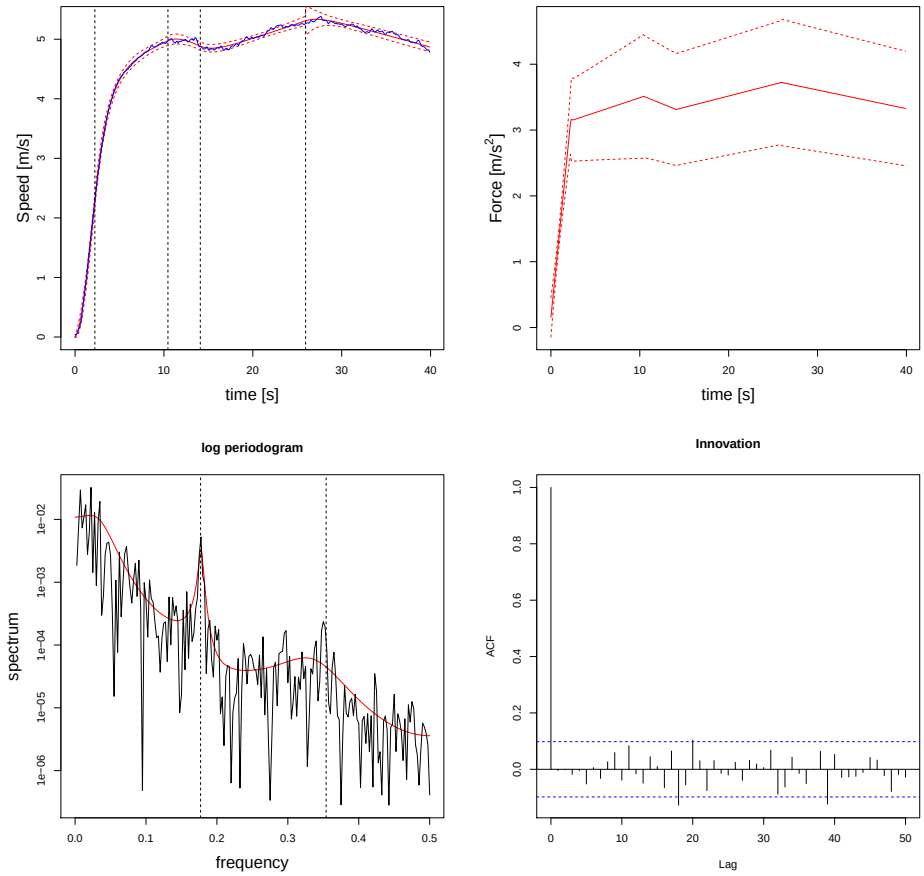

Figure 19: (Top) Simultaneous confidence interval for the fitted velocity profile and propulsive force. The model has $N_{\mathrm{ch}} = 5$ pieces. (Bottom) Diagnosis of the ARMA(6,3) model for residuals with the innovations ACF and the residual periodogram. Dotted lines correspond to the peak frequency at $\omega = 0.177$ [sample]$^{-1}$ and first harmonic at 2ω. Subject 1.

of the ODE model for $N_{\mathrm{p}} = 4$ and the ARMA(8,6) model requires fine tuning in this case, and could not been done successfully.

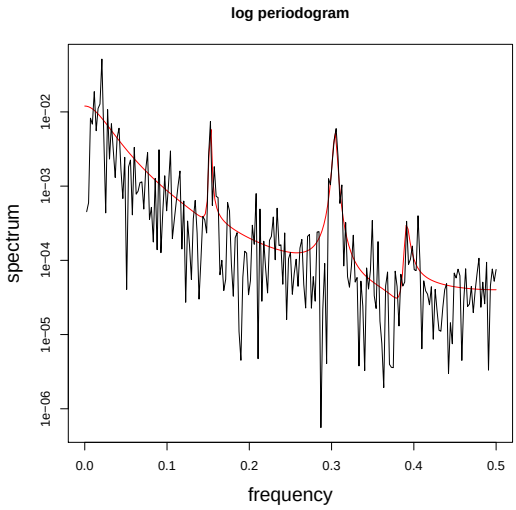

Figure 20: Periodogram of model residual, and spectral density for the fitted ARMA(8,6). Subject 2.

6 Conclusions

Discussion on ODE parameter identification This report investigated three methods of ODE parameter estimation. All of them presented coherent conclusions but with different drawbacks. The NLS method provides interesting results when accounting for correlations in residuals. In particular, it allows for comparison of embedded propulsive force models. Therefore, it also allows to investigate which portion of the velocity might be considered constant, increasing or decreasing along the race. The confidence interval in Figure 11 (Right) suggests that a smoother version of $f(t;\theta)$ could be chosen. For simplicity we worked with a piecewise linear form. Yet, the inference on the changing times η_k presents no direct interest, hence a semi-parametric structure for $f(t;\theta)$ with basis functions, e.g. B-spline basis, could be more appropriate. The propulsive force can be viewed as a spline of order 1 with variable knots, and because of the regularization property of linear ODE solutions it leads to satisfactory fits. For further research, it would be interesting to add more knots, say one at each observation time, and investigates methods to infer on the number of knots needed. In Section 2.5, we refer to the Chapter 9 of Seber and Wild (2005), that suggest information criteria or cross-validation to estimate the number of knots actually required to model data. A LASSO approach could also lead to sparse modelling of the propulsive force, while being more flexible and avoiding to fix the number of pieces in advance.

Additional flexibility with respect to the modelling can already be achieved with the GP method. We observed in Figure 17 that exploiting an intermediate value for λ and a vague model can show coherent results with the NLS method fitted with a more complex modelling. However, this gain comes at a cost. Indeed, if the value of λ is too low, the estimate for θ may become difficult to interpret. In addition, the asymptotic behavior of $\hat{\theta}(\lambda)$ is not established for finite value of λ. In particular, the covariance estimate of (11), inspired from the asymptotic distribution of $\hat{\theta}(\lambda)$ for $\lambda \to +\infty$, has ambivalent behavior. It does not account for model relaxation as it is not, in any case, a strict function of λ, but rather corresponds to the informativeness of the ODE model with respect to the fitting procedure. Besides, there still are a lot of interesting features to explore with

this method. For instance, the choice of the basis functions should be investigated. It is likely that the number of basis functions could be reduced without deteriorating the estimation. Even though it has been convenient to associate knots of B-spline basis to each time of observation, inevitably, this choice has significantly increased the time of computation. Secondly, it would be interesting to use the optimal basis choice given by the RKHS theory for the homogeneous operator $L_\theta = D + I(1/\tau)$ and analyze its relevance in this context. The analysis of generalized smoothing matrix proposed in Section 5.2 allow to gain insights into the GP method and its filtering property. The derivation of this smoothing matrix proposed in Appendix A.1 relies on the ODE's linearity, but may provide a tool to diagnos the number and the approximation quality of the basis functions.

The two-step method presents clear advantages in terms of model-fitting with respect to the previous methods. In the NLS method, the optimization procedure is fast but relies on a initial guess, and results may change with this guess. As for the GP method, the problem of finding an initial guess disappears when fitting the model with an increasing sequence for λ. This approach is computationally intensive. For each λ, we repeat the optimization, and it is slow because, at each updated value for θ, a relaxed solution is computed. The two-step method does not need an initial guess for θ, and is very fast to perform. Yet, it is important to choose an appropriate weighting function $w(t)$. Besides, the modelling of the propulsive force as function of unknown changing times is likely to be inappropriate for this method. This is due to the noisy non-parametric estimator in the phase-plane that makes the distinction between pieces difficult. If we had more time, it would be interesting to study a bootstrapping scheme to estimate the uncertainty of the parameter and compare it to the NLS method.

The modelling of the propulsive force as a piecewise linear function, with unknown changing points, could imply model regularity issues. While this have been ignored in this report, it should be investigated, in particular, for inference on changing points. An extensive discussion can be found in Chapter 9.3 of Seber and Wild (2005). The lack of regularity comes from the discontinuities of the first derivative of $E[y] = v(t;\theta)$ with respect to η_k at $t = \eta_k$. In this context the usual asymptotic theory for the changing points changes slightly, and some care is needed for model comparisons. For instance, the asymptotic distribution for likelihood ratio to test the hypothesis of no change, i.e. H_0: $\alpha_k = \alpha_{k+1}$ & $\beta_k = \beta_{k+1}$ for some $k \in \{1, ..., N_p - 1\}$, is not given by a χ_2 distribution due to the unidentifiable change point. However, the likelihood ratio's asymptotic distribution for embedded models remains as usual if all parameters are identified under the null. In particular, it means that successive pieces differ of at least one parameter to ensure identifiability of the change point under the null. The model comparisons of Section 5.1 satisfy this condition. Specific methods for testing no change between pieces are discussed in Chapter 9.2 and 9.3 of Seber and Wild (2005), and may be interesting for further research. The AIC viability in this non-regular context should also be investigated. To avoid model regularity issues one may smoothed the transition between changing points, using for instance sigmoid functions as described in Figure 12 (Left). In such context, inference for the changing points depends on the regularization parameter controlling the smoothed transition. Another way to avoid non-regularity is to model the propulsive force with an order 1 spline associated to fixed knots. The problem of inference on changing points is replaced by the problem of choosing the knots.

Perspective on applications of the runner's model For this runner's model the NLS method is probably more appropriate, as an explicit solution for the speed equation is

available. The uncertainty quantification on physiological parameters and the fitted curves are easy to obtain under reasonable hypotheses. In addition, the NLS method seems to adapt well to other runners, and offers a comprehensive approach to model velocity profile. The GP and two-step methods exploit the ODE formulation directly even if this is not essential in this context. On the other hand, the coherency of estimated parameters suggests that, for a non-linear ODE, these methods are likely to provide equivalent results while avoiding numerical integration. For the runner's modelling, these approaches are interesting for future research, in particular for the modelling of a race between two runners. For instance in Aftalion and Fiorini (2015), a wind friction term $-cv^2$ is added to the ODE equation for the velocity. This term encompasses a wind resistance. If two runners are involved this resistance is attenuated proportionally to the runners relative position. The authors show how a runner may win a race against a stronger opponent by taking advantage of wind resistance's attenuation. In addition, the ODE model of (1) could be adapted for swimming or cycling. In these sports, the wind friction is usually considered a non-linear function of the speed.

During this project, we did not have the time to perform any kind of inference on the anaerobic energy $e(t)$, nor on physiological parameters such as the $\dot{V}O_{2,max}$ $\bar{\sigma}$, or the initial anaerobic energy e_0. While these parameters are not identifiable with only speed data, it remains possible to perform simulations and look for admissible parameters given the estimated power spent $-\hat{f}(t)\hat{v}(t) = -f(t;\hat{\theta})v(t;\hat{\theta})$. Possible criteria for parameters admissibility could be a decreasing condition on $e(t)$, and a zero-energy condition at final time $e(T)$. This latter appears to be important, indeed, it is more reasonable to estimate the total anaerobic energy consumed than the anaerobic energy at its initial state.

The motivation for this project was to propose accurate interval estimation for the physiological parameters involved in the ODE of (1), in addition to a confidence interval on the fitted curve $\hat{v}(t)$. This ODE model could be used to generate an optimal velocity profile, as described in Aftaltion and Bonnans (2014). A natural tool for runners' coaching would be to compare the optimal speed profile, generated from estimated physiological performance $\hat{\theta}$, and the confidence band on the fitted curve $\hat{v}(t)$. Sports analytics have been increasingly popular during the last few years. However, the modelling, even if existent, rarely exploits measurements of the performance in the course of the effort. This is paradoxical as it seems that sport's monitoring data has become increasingly precise and available in our everyday life.

A Appendix

A.1 Derivation of the generalized smoothing matrix

This section shows the computations that lead to an explicit solution for minimization problem defined on (8). The following procedure depends in crucial way on the linearity of the ODE:

$$\dot{v}(t;\theta) = -v(t;\theta)/\tau + f(t;\theta).$$

Following the collocation approximation described in Section 3.1, the velocity is expressed in a B-spline basis system: $v(t;\theta) = \sum_{i=1}^{N_{basis}} c_i(\theta)\phi_i(t)$. Suppose an i.i.d. Gaussian behaviour of residual, and without loss of generality we can suppose that the variance satisfies $\sigma^2 = 1$. In the following, $\Phi(\cdot) = [\phi_1(\cdot), ..., \phi_{N_{basis}}(\cdot)]^T$ denotes the basis elements viewed as an element of the functional space $L^2([0,T])$ with value in $\mathbb{R}^{N_{basis}}$, i.e. $L^2([0,T])^{N_{basis}}$. The criterion of (8) becomes:

$$J(c|\theta;\lambda) = \sum_{i=1}^{N_{obs}} \left\{ y_i - c^T\Phi(t_i) \right\}^2 + \lambda \int_0^T \left\{ c^T\dot{\Phi}(t) + c^T\Phi(t)/\tau - f(t;\theta) \right\} dt$$

where y_i $i = 1, 2..., N_{obs}$ denotes the observations. We can suppose $\sigma^2 = 1$ because this optimization problem for $c \in \mathbb{R}^{N_{basis}}$ is invariant by a scaling factor. Adding σ^2 changes only the dependence on λ by a dependence on $\sigma^2\lambda$ for the final result.

Notations To express $\hat{c}(\theta)$ it is convenient to use vector calculus. We will use $\Phi_{ij} = \Phi_j(t_i)$ for the evaluation of the basis on observations time points with $i = 1, 2..., N_{obs}$ and $j = 1, 2..., N_{basis}$. We will also extent the scalar product notation in L^2 to define matrix and vector. For $i, j = 1, ..., N_{basis}$, we define $\langle g, f^T\rangle_{ij} := \int_0^T g_i(t)f_j(t)dt$ where g and f are vector fields in $\mathbb{R}^{N_{basis}}$, and $\langle g, f\rangle_i = \int_0^T g_i(t)f(t)dt$ where g is a vector field in $\mathbb{R}^{N_{basis}}$, and f is a scalar function. If both functions f and g are both scalar-valued then we have an actual scalar product $\langle f, g\rangle_{L^2([0,T])}$. A notation trick that is often used: $\langle c^T\Phi(\cdot), c^T\Phi(\cdot)\rangle_{L^2} = \langle c^T\Phi(\cdot), \Phi^T(\cdot)c\rangle_{L^2} = c^T\langle\Phi(\cdot), \Phi^T(\cdot)\rangle c$ or $\langle f, c^T\Phi(\cdot)\rangle_{L^2} = c^T\langle f, \Phi(\cdot)\rangle$, for $c \in \mathbb{R}^{N_{basis}}$ a constant. One should note that $\langle f, g^T\rangle$ and $\langle f, g\rangle$ are not actual scalar products since they are vector/matrix valued, it is a handy extension of the notations to express component-wize L^2 scalar product. In addition if f and g are vector fields $\langle f, g^T\rangle \neq \langle g, f^T\rangle$.

Explicit formula for $\hat{c}(\theta)$ Exploiting the vectorial notations introduced, and denoting the propulsive force $f_\theta := f(\cdot;\theta)$ the criterion is:

$$J(c|\theta;\lambda) = (y - \Phi c)^T(y - \Phi c) + \lambda\langle c^T\dot{\Phi}(\cdot) + \frac{c^T\Phi(\cdot)}{\tau} - f_\theta, c^T\dot{\Phi}(t) + \frac{c^T\Phi(t)}{\tau} - f_\theta\rangle_{L^2}.$$

By developing the scalar product, using the notations' properties:

$$\begin{aligned} J(c|\theta;\lambda) = (y - \Phi c)^T(y - \Phi c) + \lambda\{ & c^T[\langle\dot{\Phi}(\cdot), \dot{\Phi}(\cdot)^T\rangle + (1/\tau^2)\langle\Phi(\cdot), \Phi(\cdot)^T\rangle \\ & + 1/\tau(\langle\dot{\Phi}(\cdot), \Phi(\cdot)^T\rangle + \langle\Phi(\cdot), \dot{\Phi}(\cdot)^T\rangle)]c \\ & - 2c^T[(1/\tau)\langle f_\theta, \Phi(\cdot)\rangle + \langle f_\theta, \dot{\Phi}(\cdot)\rangle + \langle f_\theta, f_\theta\rangle_{L^2}\} \end{aligned} \quad (20)$$

A convenient way to simplify previous equation is to define the homogeneous linear operator: $L_\theta v = Dv + v/\tau$. If we define this operator component-wize on the basis

functions $\Phi(\cdot)$ by $(L_\theta\Phi(\cdot))_i = L_\theta\phi_i(\cdot)$ with $i = 1, 2..., N_{basis}$, then the criterion becomes a quadratic expression in c:

$$J(c|\theta;\lambda) = (y-\Phi c)^T(y-\Phi c) + c^T\langle L_\theta\Phi(\cdot), L_\theta\Phi(\cdot)^T\rangle c - 2c^T\langle f_\theta, L_\theta\Phi(\cdot)\rangle + \langle f_\theta, f_\theta\rangle_{L^2}. \quad (21)$$

For computations (20) enables us to note that component-wise scalar product between basis functions and its derivatives should be pre-compute in order to efficiently use any numerical minimization routine. While (21) will be used for analytical derivation. Note that $\Omega_\theta := \langle L_\theta\Phi(\cdot), L_\theta\Phi(\cdot)^T\rangle$ is a $N_{basis} \times N_{basis}$ symmetric and positive semi-definite matrix since $c^T\Omega_\theta c = ||L_\theta(c^T\Phi(\cdot))||^2_{L^2} \geq 0$. In fact, Ω_θ is positive definite because $||L_\theta(c^T\Phi(\cdot))||^2_{L^2} = 0$ implies $c^T\Phi(\cdot) \in \mathrm{Ker}(L_\theta) = \mathrm{Span}\{\exp(-t/\tau)\}$, thus $c^T\Phi(\cdot) = 0$ and $c = 0$. Indeed, none of these B-spline basis functions $\phi_1(\cdot), ..., \phi_{N_{basis}}(\cdot)$ are solution of the homogeneous ODE. Finally, the gradient of J with respect to c for a given θ is:

$$\nabla_c J(c|\theta) = -2\Phi^T(y-\Phi c) + \lambda(2\Omega_\theta c - 2b_\theta)$$

where $b_\theta := \langle f_\theta, L_\theta\Phi(\cdot)\rangle$ is a N_{basis} vector. Equalizing this equation to 0 leads to a linear equation for c. It admits an unique solution because $\Phi^T\Phi + \lambda\Omega_\theta$ is non-singular, since $\Phi^T\Phi$ and Ω_θ are positive definite.

$$\hat{c}(\theta;\lambda) = (\Phi^T\Phi + \lambda\Omega_\theta)^{-1}(\Phi^T y + \lambda b_\theta) \quad (22)$$

The Hessian matrix is straightforward to compute:

$$\frac{\partial^2 J(c|\theta)}{\partial c^T \partial c} = 2\Phi^T\Phi + 2\lambda\Omega_\theta$$

and it is a positive definite matrix. Thus $\hat{c}(\theta)$ is indeed a minimum of $J(c|\theta)$ for a fix value of θ. The uniqueness of $\hat{c}(\theta)$ follows from the convexity of J. Indeed, J is a convex function of c because it is a sum of a convex part $(v-\Phi c)^T(v-\Phi c) + \lambda c^T\Omega_\theta c$ and a linear part. The relaxed-solution of the ODE is defined by $\hat{v}(t,\theta;\lambda) = \Phi(t)\hat{c}(\theta;\lambda)$. It implies that the fitted values for the relaxed solution $\hat{v}(\theta;\lambda) := \hat{v}(\cdot,\theta;\lambda)|_{t_i} \in \mathbb{R}^{N_{obs}}$ satisfies:

$$\hat{v}(\theta;\lambda) = \Phi\hat{c}(\theta;\lambda) = \Phi(\Phi^T\Phi + \lambda\Omega_\theta)^{-1}(\Phi^T y + \lambda b_\theta).$$

The previous expression shows that the fitted values for the relaxed-solution is the sum of two contributions. The first contribution is a smoothing of the data y that operates via the smoothing matrix $H_{\theta,\lambda} = \Phi(\Phi^T\Phi + \lambda\Omega_\theta)^{-1}\Phi^T$ and depends only on τ, i.e. only on parameters defining the homogeneous ODE. The second contribution does not depend on the data y but only on model's parameters.

A.2 Sampling properties of estimator through a linearisation argument

The GP method is based on two-levels of optimization criteria. There is the inner criterion $J(c,\theta,y)$ and an outer criterion $H(c,\theta,y)$. The procedure described on Section 3.2 induces a dependence between the parameters c of the basis expansion, the structural parameters of the ODE θ and the data y. For a given value of θ, and a realization of the data $y$the minimization of $J(c|\theta,y)$ leads to $\hat{c}(\theta,y)$. Then, the minimization of $H(\hat{c}(\theta,y),\theta|y)$ with respect to θ for a given value of y leads to $\hat{\theta}(y)$. An explicit formula for $\hat{\theta}(y)$ is almost

every case impossible to obtain. To investigate the sampling properties of $\hat{\theta}(y)$, Ramsay et al. (2007) suggests the following linearisation:

$$\hat{\theta}(y) \approx \hat{\theta}(\mu) + \nabla\hat{\theta}(\mu)^T(y - \mu) \tag{23}$$

where $\mu = E(y)$. This approximation appears similar to the Delta-Method. One should note that the situation is different because the Delta-Method gives the asymptotic distribution of $g(\hat{\theta})$ when $\hat{\theta}$ is known to converge in distribution to a particular law. Previous approximation is only a first order Taylor expansion $\hat{\theta}$ as a function the data y. Suppose $\text{Cov}(y) = \Sigma$, then using that $\text{E}(\nabla\hat{\theta}(y)) = \nabla\hat{\theta}(\mu)$:

$$\text{Var}[\hat{\theta}(y)] \approx \nabla\hat{\theta}(\mu)^T\Sigma\nabla\hat{\theta}(\mu) \approx \nabla\hat{\theta}(y)^T\Sigma\nabla\hat{\theta}(y). \tag{24}$$

The last approximation holds in expectation. In order to use this approximative variance estimator, one needs to compute $\nabla\hat{\theta}(y)$. While the GP procedure does not lead to an explicit formula for $\hat{\theta}$, it is still possible to compute exactly this gradient with the implicit function theorem. For a given realization of y, the estimator $\hat{\theta}$ satisfies:

$$\frac{d}{d\theta}(H(\hat{c}(\hat{\theta}, y), \hat{\theta}, y)) = 0$$

The implicit function theorem states that it exists a neighbourhood of the data y, and such that $\frac{d}{d\theta}\left(H(\hat{c}(\hat{\theta}(z), z), \hat{\theta}(z), z)\right) = 0$ for z in this neighbourhood. To emphasize the dependence on $\hat{\theta}$, we re-define $\Gamma(z, \hat{\theta}(z)) = \frac{d}{d\theta}(H(\hat{c}(\hat{\theta}(z), z), \hat{\theta}(z), z))$. Computing the derivative with respect to z of Γ leads to:

$$\frac{d}{dz}(\Gamma(z, \hat{\theta}(z))) = 0 \Rightarrow \frac{\partial\Gamma}{\partial z}(z, \hat{\theta}(z)) + \nabla\hat{\theta}(z)\frac{\partial\Gamma}{\partial\hat{\theta}}(z, \hat{\theta}(z)) = 0$$

for z in a certain neighbourhood of y. The previous equation gives a formula for the gradient of $\hat{\theta}$ at point y:

$$\nabla\hat{\theta}(y) = -\frac{\partial\Gamma}{\partial z}(y, \hat{\theta}(y))(\frac{\partial\Gamma}{\partial\hat{\theta}}(y, \hat{\theta}(y)))^{-1}$$

where $\nabla\hat{\theta}$ has dimension $N_{obs} \times N_{par}$, $\frac{\partial\Gamma}{\partial z}$ is $N_{obs} \times N_{par}$ and $\frac{\partial\Gamma}{\partial\hat{\theta}}$ is $N_{par} \times N_{par}$. The notation for partial derivatives of Γ hides details about the derivatives of $\frac{d}{d\theta}(H(\hat{c}(\hat{\theta}(z), z), \hat{\theta}(z), z))$. Indeed, the partial derivative of Γ with respect to z induces a total derivative of H with respect to z but considering $\hat{\theta}(z)$ as a constant. The intermediate function $\Gamma(z, \theta(z))$ has the following expression:

$$\Gamma(z, \hat{\theta}(z)) = \frac{\partial}{\partial\hat{\theta}}\hat{c}(\hat{\theta}(z), z)\frac{\partial}{\partial\hat{c}}H(\hat{c}(\hat{\theta}(z), z), \hat{\theta}(z), z) + \frac{\partial}{\partial\hat{\theta}}H(\hat{c}(\hat{\theta}(z), z), \hat{\theta}(z), z).$$

The second term is an *explicit* derivative of H with respect to θ considering that $\hat{c}$ does not depend on θ. The derivatives of Γ with respect to z and θ are given below. For the purpose of clarity the functions evaluation point are omitted.

- $\frac{\partial}{\partial z}\Gamma$, or the full derivative with respect to z of $\frac{d}{d\theta}H(\hat{c}(\hat{\theta}(z), z), \hat{\theta}(z), z)$ not considering dependence of $\hat{\theta}$ on z:

$$\frac{\partial}{\partial z}\Gamma = \frac{\partial^2\hat{c}}{\partial\hat{\theta}\partial z}\frac{\partial H}{\partial\hat{c}} + \frac{\partial^2 H}{\partial c\partial z}\frac{\partial\hat{c}}{\partial\hat{\theta}}^T + \frac{\partial\hat{c}}{\partial z}\frac{\partial^2 H}{\partial c\partial c}\frac{\partial\hat{c}}{\partial\hat{\theta}} + \frac{\partial\hat{c}}{\partial z}\frac{\partial^2 H}{\partial\hat{\theta}\partial c} + \frac{\partial^2 H}{\partial\hat{\theta}\partial z}$$

- $\frac{\partial}{\partial\hat{\theta}}\Gamma$, or the full derivative of $\frac{d}{d\hat{\theta}}H(\hat{c}(\hat{\theta}(z),z),\hat{\theta}(z),z)$ with respect to θ

$$\frac{\partial}{\partial\hat{\theta}}\Gamma = \frac{\partial^2\hat{c}}{\partial\hat{\theta}\partial\hat{\theta}}\frac{\partial H}{\partial\hat{c}} + 2\frac{\partial\hat{c}}{\partial\hat{\theta}}\frac{\partial^2 H}{\partial\hat{c}\partial\hat{\theta}}^T + \frac{\partial\hat{c}}{\partial\hat{\theta}}\frac{\partial^2 H}{\partial\hat{c}\partial\hat{c}}\frac{\partial\hat{c}}{\partial\hat{\theta}}^T + \frac{\partial^2 H}{\partial\hat{\theta}\partial\hat{\theta}}$$

  Fortunately, these derivatives are computed for a general ODE model in the `CollocInfer` package.

References

J. O. Ramsay, G. Hooker, D. Campbell and J. Cao, *Parameter Estimation for Differential Equations: A Generalized Smoothing Approach.* Journal of the Royal Statistical Society: Series B, Volume 69, Issue 5, pp. 741–796, 2007

X. Qui and H. Zhao, *Asymptotic efficiency and finite-sample properties of the generalized profiling estimation of parameters in ordinary differential equations*, Annals of Statistics, Vol. 38, No. 1, pp. 435-481, 2010

G.A.F. Seber and C.J. Wild, *Nonlinear Regression*, Wiley, 2005

A. Aftalion and F. Bonnans, *Optimization of running strategies based on anaerobic energy and variations of velocity.* SIAM Journal of Applied Mathematics, 74 (5), pp. 1615-1636, 2014

S. Gsteiger , F. Bretz and W. Liu *Simultaneous Confidence Bands for Nonlinear Regression Models with Application to Population Pharmacokinetic Analyses*, Journal of Biopharmaceutical Statistics, 21:4, pp. 708-725, 2011

N. Heckman *The theory and application of penalized methods or Reproducing Kernel Hilbert Spaces made easy*, arXiv:1111.1915, 2011

D.A. Campbell and O. Chkrebtii, *Maximum profile likelihood estimation of differential equation parameters through model based smoothing state estimates*, Mathematical Biosciences Volume 246, Issue 2, pp. 283–292, 2013

N. Brunel, *Parameter estimation of ODE's via nonparametric estimators.* Electronic Journal of Statistics, Vol. 2 (2008), pp. 1242-1267, 2009

J. B. Keller, *Optimal velocity in a race.* Amer. Math. Monthly, pp. 474-480, 1974

C. Hanon and C. Thomas, *Effects of optimal pacing strategies for 400-, 800-, and 1500-m races on the [Vdot]O2 response*, Journal of Sports Sciences, 29:9, pp. 905-912, 2011

A. B. Pitcher, *Optimal Strategies for a Two-Runner Model of Middle-Distance Running.*, SIAM Journal on Applied Mathematics, 70(4), pp. 1032-1046, 2009

A. Aftalion, and C. Fiorini, *A two-runners model: optimization of running strategies according to the physiological parameters.*, arXiv:1508.00523v2, 2015

A. Aftalion, L-H. Despaigne, A. Frentz, P. Gabet, A. Lajouanie, M-A. Lorthiois, L. Roquette and C. Vernet *How to identify the physiological parameters and run the optimal race* MathS In Action, 7 no. 1, pp. 1-10, 2016

Y. Sun and G. Genton, *Functionnal Boxplots*, Journal of Computational and Graphical Statistics, Volume 20, Issue 2, pp. 316-334, 2012