
ON THE PARAMETERS ESTIMATION OF HIV DYNAMIC MODELS *

Authors: DIANA ROCHA

– Center for R&D in Mathematics and Applications (CIDMA),
University of Aveiro, Portugal
`diana.isa.rocha@ua.pt`

SÓNIA GOUVEIA

– Institute of Electronics and Informatics Engineering of Aveiro (IEETA)
and CIDMA, University of Aveiro
`sonia.gouveia@ua.pt`

CARLA PINTO

– Centre of Mathematics of the University of Porto (CMUP) and
School of Engineering, Polytechnic of Porto
`cpinto@fc.up.pt`

MANUEL SCOTTO

– Center for Computational and Stochastic Mathematics (CEMAT) and
Department of Mathematics, IST, University of Lisbon
`manuel.scotto@tecnico.ulisboa.pt`

JOÃO NUNO TAVARES

– Centre of Mathematics of the University of Porto (CMUP)
`jntavar@fc.up.pt`

EMÍLIA VALADAS AND LUÍS FILIPE CALDEIRA[†]

– Infectious Disease Service, Hospital Santa Maria, Lisbon (HSM/CHLN)
`evaladas@medicina.ulisboa.pt` and
`luis.caldeira@chln.min-saude.pt`

Received: October 2018

Revised: January 2019

Accepted: March 2019

Abstract:

- This work proposes an estimation method to obtain the optimal parameter estimates of a mathematical model, from a set of CD4⁺T values collected in a HIV patient. To this end, the following scheme is adopted: the first step consists in selecting an initial estimate for the model's parameters as that having minimum square error, from a set of uniform randomly generated candidates. In the second step, the initial solution is refined by an optimization algorithm with constraints and bounds (imposed by physiology), resulting on the optimal estimate. The proposed method is validated through a simulation study and illustrated with an application to a real data set of CD4⁺T cells counts for several HIV patients.

*The opinions expressed in this text are those of the authors and do not necessarily reflect the views of any organization.

[†] Deceased January 29, 2019.

Key-Words:

- *Parameter estimation; nonlinear programming; mathematical models; human immunodeficiency virus (HIV).*

AMS Subject Classification:

- 49A05, 78B26.

1. INTRODUCTION

In the clinical follow-up of a HIV/AIDS patient, the viral load values and the CD4⁺T cells count, observed over time, constitute a set of non-equally spaced observations. In general, no information between consultations is available. In this context, it is clinically relevant to develop methods able to obtain a more complete description of the individual time evolution, either between consecutive consultations or for the prediction of evolution or disease progression. Viral dynamic models can be formulated through a system of nonlinear ordinary differential equations, which enables to describe the temporal evolution of the clinical parameters of a HIV patient [2, 10, 11]. In the last decades, a few literature studies show applications and developments in statistical methodologies for model inference including those based on Bayesian inference [6]. Briefly, the Bayesian approach incorporates non-informative prior distributions and yet the corresponding algorithms require initial estimates for model's parameters in order to carry out the iterative updates of the parameters. For the estimation of this initial estimates, the most commonly used approaches in practice are based on nonlinear least squares [7, 8]. In this context, this work presents a nonlinear programming approach to obtain the optimal estimates for the parameters of a HIV dynamic model. Our proposal differs from previous approaches in the fact that we add restrictions on the optimal estimate so that it verifies an equal contribution of negative and positive deviations from observations. Furthermore, the optimal estimate is restricted to be in-between lower and upper physiological bounds. Note that the least square methods are implemented as optimization problems requiring initial solutions to start the iterations. To cope with this limitation, we consider as initial solution the minimum square error solution from a set of 1000 uniform randomly-generated candidates on a uniform distribution delimited by the lower and upper physiological bounds. Therefore, the proposed method is fully automatic and does not require any other information to provide the optimal estimate of the model's parameters besides the data.

This paper is organized as follows: the methods concerning the description of the mathematical HIV model and the estimation approach developed to obtain the initial conditions for the model's parameters are presented in Section 2. The estimation approach is illustrated with simulated data that mimics the individual temporal trajectories of three HIV patients. The simulation strategy is described in Section 3, whereas the results on simulation and on real data from six HIV patients [12] are presented in Section 4. The selected patients were chosen according to some conditions, namely having started an antiretroviral treatment at the beginning of the trial [12]. Finally, Section 5 is devoted to conclusions.

2. METHODS

The dynamics of the HIV/AIDS infection is described through the mathematical model presented in Section 2.1. Model's parameters are estimated from the nonlinear programming approach described in Section 2.2. The developed algorithms and other software code used in this work were implemented in MATLABTM (version R2015a), The Mathworks Inc., MA, USA.

2.1. Mathematical model

The mathematical model considered in this work translates known physiological relationships between viral load and CD4⁺T cells and incorporates parameters having clinical interpretation. We consider a modified version of the mathematical model in Stafford *et al.* [14] for the dynamics of HIV/AIDS infection, including an additional parameter ϵ that denotes the effectiveness of the antiretroviral therapy [9]. The model is represented as

$$(2.1) \quad \begin{aligned} \frac{dT(t)}{dt} &= \lambda - d_1 T(t) - (1 - \epsilon)k_1 T(t)V(t), \\ \frac{dT^*(t)}{dt} &= (1 - \epsilon)k_1 T(t)V(t) - \delta T^*(t), \\ \frac{dV(t)}{dt} &= \pi_1 T^*(t) - cV(t), \end{aligned}$$

where the state variables are the viral load $V(t)$ and the number of CD4⁺T cells defined as $CD4(t) = T(t) + T^*(t)$, with $T(t)$ and $T^*(t)$ representing the number of uninfected and infected CD4⁺T cells, respectively. Furthermore, for simplicity in notation we denote $(T(0), T^*(0), V(0)) = (T_0, T_0^*, V_0)$ as the initial condition of the model. Along with the states variables, the mathematical model also incorporates parameters with clinical interpretation, namely $\theta = (d_1, \epsilon, k_1, \delta, \pi_1, c)$, with definition and units listed in Table 1.

The mathematical model in (2.1) can be alternatively defined from the flow-chart displayed in Figure 1.

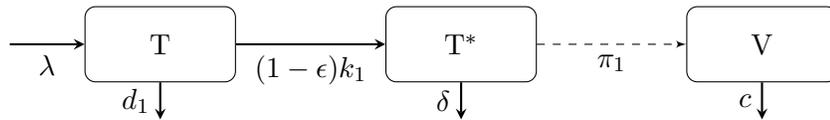


Figure 1: Schematic diagram of the model (2.1).

The chart presents a compartmental description of the model that translates the evolution of the disease at the patient level. Within each compartment

Parameter	Definition	Units
d_1	difference between rate loss from cell death and rate gain due to cell division	day ⁻¹
$\lambda = T_0 d_1$	proliferation rate of uninfected target cells	cells ml ⁻¹ day ⁻¹
ϵ	effectiveness of therapy	
k_1	infectivity rate	ml day ⁻¹
δ	death rate of infected cells	day ⁻¹
π_1	average number of virions produced by a single infected cell	day ⁻¹
c	clearance rate of free virions	day ⁻¹

Table 1: Definition and units of the parameters included in (2.1).

there are CD4⁺T cells (non-infected or infected) or viral load. In this representation, these units can move between compartments. For instance, the susceptible CD4⁺T cells in compartment T move to compartment T^* (infected cells) after being infected with HIV at a rate equal to $(1 - \epsilon)k_1$, and infected CD4⁺T cells of compartment T^* die at rate δ .

2.2. Nonlinear programming

The parameters in θ can be estimated from a set of $CD4(t)$ values collected in one HIV patient at its clinical follow-up appointments over time. Let $CD4(t_i)$ be the observed number of CD4⁺T cells at time $t_i, i = 1, 2, \dots, n$. Furthermore, define $\widehat{CD4}(t_i) = T(t_i) + T^*(t_i)$ as the estimate of $CD4(t_i)$ provided by the mathematical model (2.1). The *optimal* parameter estimates, say $\hat{\theta}$, can be obtained by minimizing the square error between model estimates and observed CD4 values. In accordance with other literature studies [6], we considered a log10-transformation on the parameters to ensure their positiveness and to stabilize the $CD4(t)$ variance. Thus, the nonlinear programming algorithm can be formulated as

$$\begin{aligned}
 & \text{minimize} && f(\theta) = \sum_{i=1}^n (\widehat{CD4}(t_i) - CD4(t_i))^2 = \sum_{i=1}^n e_{t_i}^2 \\
 (2.2) \quad & \text{subject to} && \sum_{i=1}^n e_{t_i} = 0 \\
 & \text{and} && \mathbf{lb} \leq \theta \leq \mathbf{ub}
 \end{aligned}$$

where the restriction guarantees that $\hat{\theta}$ verifies equal contribution of negative and positive deviations from observations. Also, $\hat{\theta}$ is restricted to physiological lower and upper bounds, respectively $\mathbf{lb} = (0.01, 0, 10^{-11}, 0.24, 50, 2.39)$ and $\mathbf{ub} = (0.02, 1, 10^{-5}, 0.7, 10000, 23)$ [5, 14]. This optimization procedure was implemented with the MATLABTM function *fmincon*, that starts at an initial solution

θ^* to find a minimizer $\hat{\theta}$ of $f(\theta)$ subject to the above-mentioned restrictions and bounds. The initial solution θ^* was obtained as that minimizing $f(\theta)$ in a set of 1000 candidates randomly generated from a multivariate uniform distribution on \mathbf{lb} and \mathbf{ub} .

The HIV dynamic model (2.1) was implemented with MATLABTM function *ode45*. This function makes use of an explicit Runge-Kutta formula, namely the Dormand-Prince pair [4], that computes the solution at time t_k based on the solution at time t_{k-1} . Furthermore, when the integration is considered in a time span, the algorithm runs with a variable time step for efficient computation. In this case, temporal resampling is needed to obtain the solutions at specific $t_i, i = 1, 2, \dots, n$ (continuous time). Alternatively, the solver can provide the solution at requested time points t_i with its own built-in interpolation algorithm (discrete time). The differences between continuous/discrete time solutions were used to determine if differences between solutions evaluated at θ and at $\hat{\theta}$ are numerically relevant.

3. SIMULATED DATA

The estimation procedure described above is illustrated through a simulation study. In this work, regularly spaced $CD4(t)$ and $V(t)$ observations are obtained within the interval $[0, 120]$ (days), by numerical Runge-Kutta integration of Equation (2.1). Note that simulating data for regularly spaced observations is not a limitation, as the model (2.1) can, in the same way, be applied to obtain non-equally spaced measurements. We reproduce the evolution of three HIV patients with parameters θ_0 presented in Table 2 [14]. Moreover, we considered the initial conditions $(T_0, T_0^*, V_0) = (11 \times 10^3, 0, 10^{-6})$ with units $(\frac{cells}{ml}, \frac{cells}{ml}, \frac{copies}{ml})$, respectively, that mimics a condition with a large initial number of uninfected cells T_0 and low values for the initial number of infected cells T_0^* and viral load V_0 .

Patient	d_1	k_1	δ	π_1	c
1	0.013	0.46×10^{-6}	0.40	980	3
2	0.012	0.75×10^{-6}	0.39	790	3
3	0.017	0.80×10^{-6}	0.31	730	3

Table 2: Parameter values used for the simulation of 100 replicas for 3 patients [14], in a total of 300 simulations. $\epsilon = 0$ is considered. Further description of the parameters can be found in Table 1.

Within this setting, we obtain a set of $n = 18$ observations representing the temporal trajectory of each patient in a clinical follow-up every 7 days ($t_i \in \{0, 7, 14, 21, 28, \dots, 119\}, i = 1, 2, \dots, 18$ and $t_1 = 0$ is the time instant of the first $CD4^+$ T observation of the patient). Afterwards, 300 replicas (100 replicas for each patient) of that trajectory are randomly generated, by adding an error

to the CD4⁺T values, in accordance with the fact that laboratory CD4⁺T measurements have an error of about 20% of the measured value (i.e. $e \sim N(0, \sigma_e^2)$) [15]. Note that the quadratic deviation (of the realizations) of e from zero is such that $\sum_{i=1}^n e_{t_i}^2 = f(\theta_0) \approx \sigma_e^2(n - 1)$, as θ_0 is the simulation reference. For each replica, we obtain $\hat{\theta}_0$ as the solution of the optimization problem. For the purpose of illustration, Figure 2 shows one replica of each patient and highlights the similarities and differences between patients.

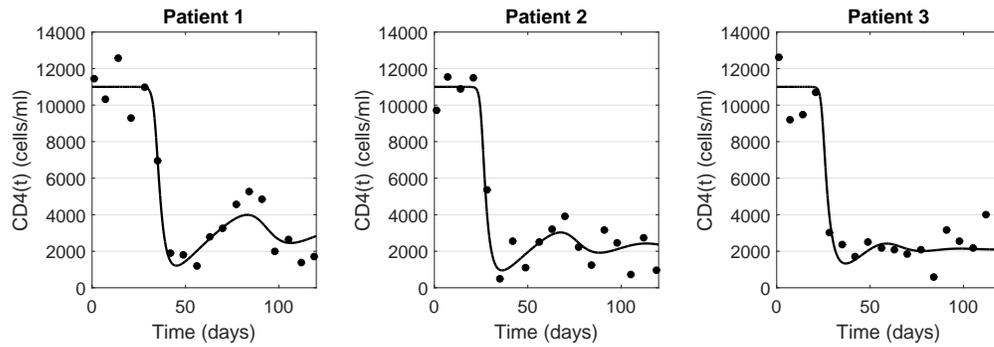


Figure 2: $CD4(t)$ trajectory over time for the reference patients with the θ parameters in Table 2. The circles represent the observations obtained for one replica of that patient [13].

After infection with HIV, there is an acute phase characterized by an accentuated decay of the number of CD4⁺T cells, since they are HIV preferred target. This can be observed, in the graphs, between 30 and 40 days approximately. The immune system tries to fight the virus by producing antibodies. After this phase, the chronic phase of infection starts, defined by the body recovery. It is observed a slight increase in the number of CD4⁺T cells. This feature is shared for the three patients although the minimum and the maximum values of the CD4⁺T cells vary between patients, before the CD4⁺T cells reach an almost constant value. Biologically, since the CD4⁺T cells play a key role in the immune response to pathogens, the differences in those values (namely, the minima) may induce the development of more severe infections, e.g., certain types of cancers and non-AIDS diseases.

4. RESULTS

In this section, the results are detailed for the simulations produced for patient 2 (see corresponding set of reference parameters θ_0 in Table 2). The performance evaluation of the model with respect to simulated data was assessed by $f(\theta)$ either appraised for θ_0 (the reference simulation parameters) or $\hat{\theta}_0$ (the parameters estimated from simulated data). The function $f(\theta)$ translates the goodness-of-fit of the model-based observations with respect to simulated data

(see equation 2.2). Figure 3(a) shows the distribution of $f(\boldsymbol{\theta}_0)$ for the 100 replicas of patient 2, where it is possible to observe that the values obtained for all replicas are centered around that evaluated for the reference parameters in $\boldsymbol{\theta}_0$. Figures 3(b–c) display the CD4 trajectory lines obtained from $\boldsymbol{\theta}_0$ and from $\hat{\boldsymbol{\theta}}_0$ for two replicas with $f(\boldsymbol{\theta}_0)$ close to and higher than the reference value $f(\hat{\boldsymbol{\theta}}_0)$, respectively. As is illustrated in Figure 3(b), the estimation procedure provided similar curves for $f(\boldsymbol{\theta}_0)$ close to $f(\hat{\boldsymbol{\theta}}_0)$. Moreover, as presented in Figure 3(c) for a replica with $f(\boldsymbol{\theta}_0)$ higher than $f(\hat{\boldsymbol{\theta}}_0)$ ($f(\boldsymbol{\theta}_0) = 1.6 \times 10^7$ and $f(\hat{\boldsymbol{\theta}}_0) = 1.0 \times 10^7$), there is a relevant improvement of fit from $\boldsymbol{\theta}_0$ to $\hat{\boldsymbol{\theta}}_0$, as $\hat{\boldsymbol{\theta}}_0$ produces a curve which is clearly more adjusted to the simulated data than that obtained with $\boldsymbol{\theta}_0$. Figure 3(c) also suggests that the observations do not contribute equally to the performance increase e.g. residuals at high derivative values (black dots) are increased for $f(\boldsymbol{\theta}_0)$ and reduced when $\boldsymbol{\theta}_0$ is replaced by $\hat{\boldsymbol{\theta}}_0$.

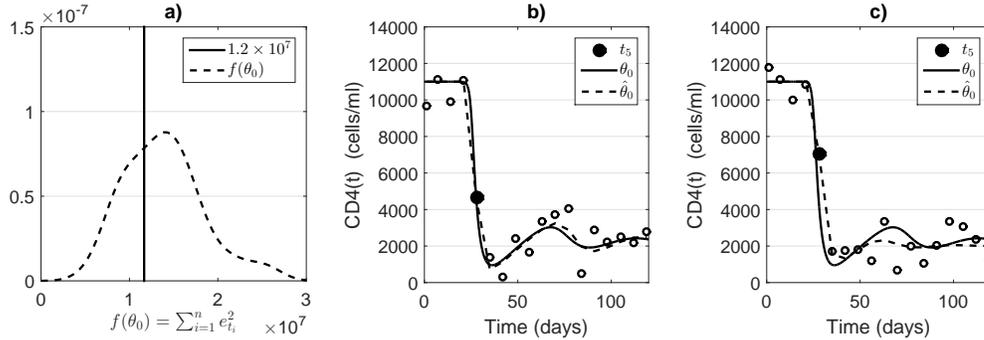


Figure 3: (a) Distribution of $f(\boldsymbol{\theta}_0)$ evaluated for 100 replicas of patient 2 (i.e. $\mathbf{s}_{\hat{\boldsymbol{\theta}}_0}^2(n-1)$ where $\hat{\boldsymbol{\theta}}_0$ are the residuals of the model with parameters $\boldsymbol{\theta}_0$, for each replica). The vertical line locates $\sigma_{\hat{\boldsymbol{\theta}}_0}^2(n-1) = 1.2 \times 10^7$ used in the simulation. (b–c) CD4 trajectory line from $\boldsymbol{\theta}_0$ and optimized $\hat{\boldsymbol{\theta}}_0$ for two different replicas: (b) $f(\boldsymbol{\theta}_0) = f(\hat{\boldsymbol{\theta}}_0) = 1.2 \times 10^7$ and (c) $f(\boldsymbol{\theta}_0) = 1.6 \times 10^7$ and $f(\hat{\boldsymbol{\theta}}_0) = 1.0 \times 10^7$. The circles represent the simulated observations and the black dot highlights time t_5 .

Figure 4 further compares the modeling results for the replicas for patient 2. As observed in Figures 4(a–b), the distribution of $f(\hat{\boldsymbol{\theta}}_0)$ is more shifted towards the small deviations than $f(\boldsymbol{\theta}_0)$ and $f(\boldsymbol{\theta}_0) - f(\hat{\boldsymbol{\theta}}_0)$ is positive for almost all replicas, thus evidencing that lower squared errors are achieved for $\hat{\boldsymbol{\theta}}_0$. Moreover, as illustrated in Figures 4(c–d), the $f(\boldsymbol{\theta}_0) - f(\hat{\boldsymbol{\theta}}_0)$ differences become higher than those obtained by choosing continuous/discrete time option for the model numerical resolution. This suggests that differences between $f(\boldsymbol{\theta}_0)$ and $f(\hat{\boldsymbol{\theta}}_0)$ are indeed relevant.

The result illustrated in Figure 3(c) suggested that the observations do not contribute equally to the performance increase with special emphasis on high derivative CD4 values. Figure 5(a) shows the association between performance increase of $\hat{\boldsymbol{\theta}}_0$ with respect to $\boldsymbol{\theta}_0$, as measured by $f(\boldsymbol{\theta}_0) - f(\hat{\boldsymbol{\theta}}_0)$, and the dispersion

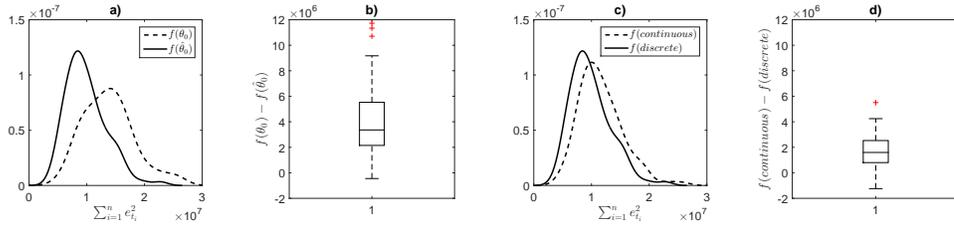


Figure 4: (a) Distribution of $f(\theta_0)$ and $f(\hat{\theta}_0)$ for 100 replicas of patient 2. (b) Boxplot of the paired differences $f(\theta_0) - f(\hat{\theta}_0)$. (c-d) Same representation as (a-b) for $f(\hat{\theta}_0)$ and continuous/discrete time.

of the residuals introduced in the simulation process. The correlation turns out to be moderate for this patient ($r = 0.60$). The effect of the residual at each time t_i was further investigated, by computing the correlation between $f(\theta_0) - f(\hat{\theta}_0)$ and the squared residual value at time t_i . Figure 5(b) shows a high correlation between $e_{t_5}^2$ and performance increase ($r = 0.91$), where higher $e_{t_5}^2$ values are associated with higher performance improvement. Furthermore, note that the large part of the residuals dispersion is due to the contribution of e_{t_5} . This analysis corroborates that the observations do not contribute equally to the performance increase. In this case, t_5 corresponds to the time point with the largest residual values for θ_0 and highest derivate in the CD4 curve (Figures 3(b-c)).

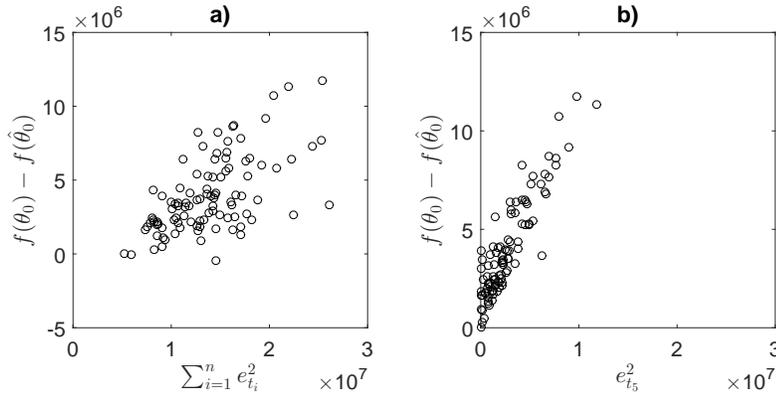


Figure 5: Dispersion diagram of $f(\theta_0) - f(\hat{\theta}_0)$ as a function of a) $\sum_{i=1}^n e_{t_i}^2$ and b) $e_{t_5}^2$, the time that maximizes correlation between $f(\theta_0) - f(\hat{\theta}_0)$ and $t_i, i = 1, 2, \dots, n$. For the remaining time points the absolute correlation was < 0.20 . Each dot represents one of the 100 replicas for patient 2.

An overall comparison of the 100 replicas simulated for the 3 patients (in a total of 300 replicas) is presented in Figure 6. As observed in Figure 6(a), the distribution of $f(\hat{\theta}_0)$ is more shifted towards the small deviations than $f(\theta_0)$ for all patients such that $f(\theta_0) - f(\hat{\theta}_0)$ is positive for almost all replicas. Again, that lower squared errors are achieved for $\hat{\theta}_0$ for all patients. Moreover, as illustrated

in Figure 6(b), the differences $f(\theta_0) - f(\hat{\theta}_0)$ are higher than those obtained by choosing continuous/discrete time option on the model numerical resolution. This suggests that differences between $f(\theta_0)$ and $f(\hat{\theta}_0)$ are indeed relevant regardless of the simulated patient.

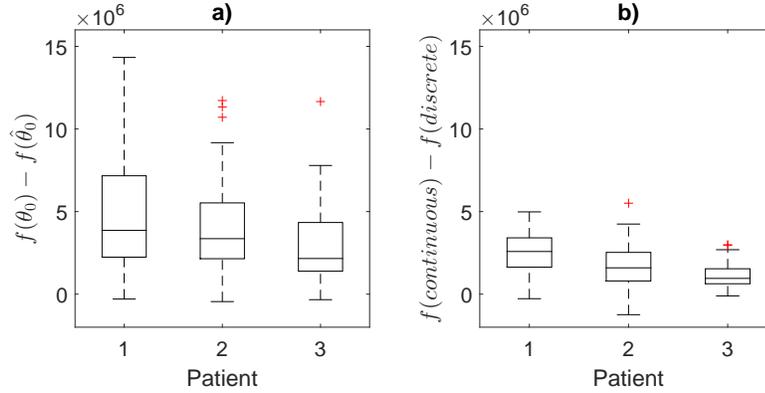


Figure 6: Boxplots of the paired differences (a) $f(\theta_0) - f(\hat{\theta}_0)$ (b) $f(\hat{\theta}_0)$ and continuous/discrete time, for each patient.

Finally, the contribution of the different observations to the performance increase is shown in Figure 7 for the 3 patients. Again, the results point out that the correlation between $f(\theta_0) - f(\hat{\theta}_0)$ and the squared residual value at a given time t_i is highest for the time point with the largest residual values for θ_0 and highest derivate in the CD4 curve. The maximum correlation between these variables reach 0.86 for patient 1 and 0.90 for patient 3. For the remaining time points, the absolute correlation is lower than 0.2 for all patients.

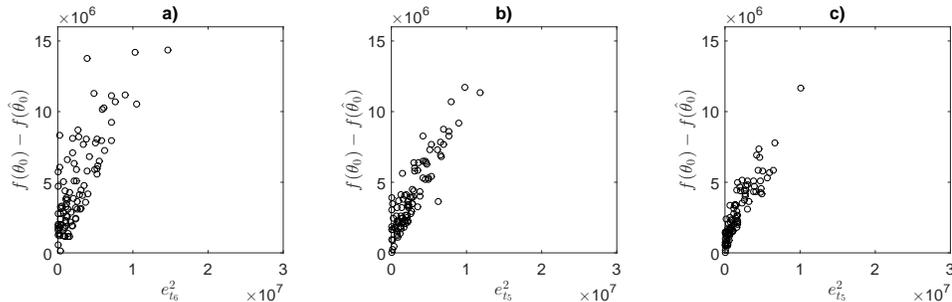


Figure 7: Dispersion diagram of $f(\theta_0) - f(\hat{\theta}_0)$ as a function of the squared error evaluated for the time that maximizes correlation between these variables. Data for patient (a) 1, (b) 2 and (c) 3.

In this work, the methods were also applied to a real data set of CD4⁺T cells count from six HIV patients [12]. The patients involved in the trial were chosen according to some conditions, namely being infected either by HIV-1 or HIV-2 type virus, being naive of any treatment at the beginning of the trial, not having hepatitis B or C virus co-infections during the 6 months before the inclu-

sion into the trial and having started an antiretroviral treatment at the beginning of the trial (therefore, we considered $\epsilon > 0$ in equation 2.1). The follow-up of the CD4⁺T cells data for all patients can be found in the study of Rivadeneira et al. [12]. Table 3 presents the follow-up of Patient 6.

Day	0	1	2	4	11	21	31	44	59	74	92	199
CD4	405	403	480	522	510	436	479	661	615	445	686	716

Table 3: CD4⁺T cells count ($\frac{cells}{mm^3}$) per time of observation (days) for Patient 6 [12]. The data for the remaining patients used in this work correspond to Patients 1, 2, 3, 4 and 8 and can be obtained in the paper of Rivadeneira and colleagues [12].

Since these patients start an antiretroviral treatment at the beginning of the trial, it is necessary to estimate the time of the infection. To this end, we start the optimization from the minimum value of $\widehat{CD4}(t_i) = T(t_i) + T^*(t_i)$, where $\widehat{CD4}(t_i)$ is the estimate of $CD4(t_i)$ provided by the mathematical model (2.1) at time $t_i, i = 1, 2, \dots, n$. Moreover, we also need to estimate the initial value T_0 for that specific patient. This can be done by estimating additionally the parameter T_0 (besides the vector of parameters $\hat{\theta}$) with lower and upper bounds given by $lb = 100$ and $ub = 1200000$, respectively [3].

Figure 8(a) shows the estimated trajectory for Patient 6, with units' $\frac{cells}{ml}$ ($\frac{cells}{mm^3} = \frac{cells}{ml} \times 10^3$), obtained from the optimal estimates $\hat{\theta}_6 = (0.013, 0.908, 3.2 \times 10^{-9}, 0.693, 9999.999, 2.390)$ and initial number of uninfected cells $\widehat{T}_0^6 = 787728 \frac{cells}{ml}$. The results indicate that the effectiveness of therapy is approximately 91% ($\hat{\epsilon}_6 = 0.908$) and that the infected cells die at a rate of 0.693 per day ($\hat{\delta}_6 = 0.693$). Also, the analysis of the curve suggests that Patient 6 was infected around 68 days before being included in the trial ($t = 0$). Figure 8 (b) resumes the results obtained for Patients 1, 2, 3, 4 and 8 [12] and illustrates the inter-subject variability of HIV individual patterns before and during antiretroviral treatment. The effectiveness of therapy $\hat{\epsilon}$ is above 90% for all patients whereas the daily death rate of infected cells $\hat{\delta}$ varies between 0.24 and 0.69. The analysis of the curves provides an estimate time of infection of approximately 28 days for Patient 2, 49 days for Patients 4 and 8 and 55 days for the remaining patients. Finally, \widehat{T}_0 varies between 237902 (Patient 3) and 980902 (Patient 2). In all cases, $f(\hat{\theta})$ varies between 0.001×10^7 and 0.021×10^7 which is lower than that observed in the simulated data. Such result is expected because in real data there are no points of very large residuals like in the simulation condition (e.g. t_5 or t_6 in the simulation condition, depending of the patient). Therefore, the results suggest that methods' performance in real data is not worse than that in simulated replicas of the same patient. This is an important result because the simulated data is drawn from the mathematical model, on the contrary of the real data, and thus good performances in terms of goodness-of-fit are expected for the curves estimated from the simulated data. Note that the order of magnitude of Figure 8 is different from that of Figure 2, since CD4 values vary between 100 and 1200000 [3]. Thus, we conclude that the patients of the simulation are worse

off than the patients in the study of Rivadeneira et al. [12].

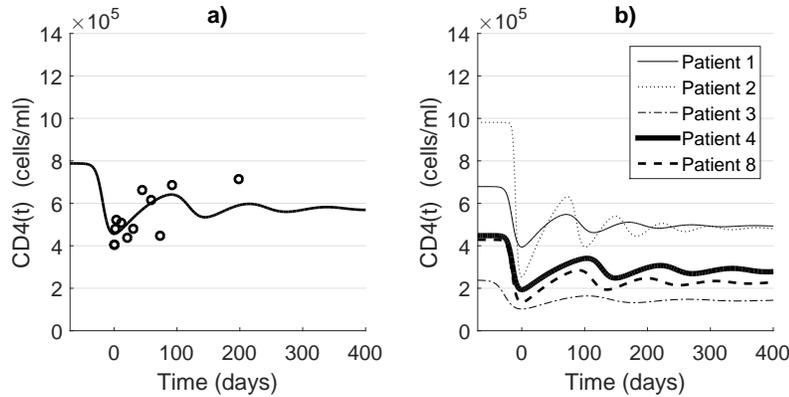


Figure 8: (a) $CD4(t)$ trajectory over time for Patient 6 with the $\hat{\theta}_6$ parameters. The circles represent the observations for Patient 6 in Table 3. (b) Same representation for Patients 1, 2, 3, 4 and 8 [12].

5. CONCLUSION

This work addresses the problem of estimating the parameters of a HIV dynamic model from a set of observations. Our method considers the minimization of the square error between model estimates and observed CD4 values, with a restriction that guarantees that the optimal solution $\hat{\theta}_0$ verifies equal contribution of negative and positive deviations from observations. Furthermore, the $\hat{\theta}_0$ estimates are restricted to lower and upper physiological bounds, which allows us to obtain a fully automatic method, in which it is not necessary to introduce an initial condition. The proposed method is validated via a data simulated with reference parameters θ_0 to mimic 3 different patients. The results indicate that the replacement of θ_0 by $\hat{\theta}_0$ decreases the fit error in a value that is greater than the difference between the fit errors obtained in the continuous and in the discrete options on the model numerical resolution. Therefore, the performance increase when replacing θ_0 by $\hat{\theta}_0$ is numerically relevant. Finally, the algorithm provides adequate $\hat{\theta}_0$ estimates (i.e. with low fit error to simulated and to real data), which enables a proper characterization of the temporal trajectory of a HIV patient.

ACKNOWLEDGMENTS

This work was partially funded by the Foundation for Science and Technology, FCT, through national (MEC) and European structural (FEDER) funds,

in the scope of UID/MAT/04106/2019 (CIDMA/UA), UID/CEC/00127/2019 (IEETA/UA), UID/Multi/04621/2019 (CEMAT/UL) and UID/MAT/00144/2019 (CMUP/UP) projects. Diana Rocha acknowledges the FCT grant with reference SFRH/BD/107889/2015.

The revision of this paper has been overshadowed by Dr. Caldeira's departure on January 2019. As the Infectious Disease Service Director, he always welcomed our scientific collaboration with great enthusiasm. In sorrow, we dedicate this work to his memory.

REFERENCES

- [1] BOFILL, M.; JANOSSY, G.; LEE, C.A.; MACDONALD-BURNS, D.; PHILLIPS, A.N.; SABIN C.; TIMMS A.; JOHNSON, M.A. and KERNOFF, P.B. (1992). Laboratory control values for CD4 and CD8 T lymphocytes. Implications for HIV-1 diagnosis, *Clinical & Experimental Immunology*, **88**, 2, 243–252.
- [2] BONHOEFFER, S.; MAY, R.M.; SHAW, G.M. and NOWAK, M.A. (1997). Virus dynamics and drug therapy, *Proceedings of the National Academy of Sciences of the United States of America*, **94**, 6971–6976.
- [3] CONWAY, J.M. and PERELSON, A.S. (2015). Post-treatment control of HIV infection, *Proceedings of the National Academy of Sciences of the United States of America*, **112**, 17, 5467–5472.
- [4] DORMAND, J.R. and PRINCE, P.J. (1980). A family of embedded Runge-Kutta formulae, *Journal of Computational and Applied Mathematics*, **6**, 19–26.
- [5] HADJIANDREOU, M.M.; CONEJEROS, R. and WILSON, D.I. (2009). Long-term HIV dynamics subject to continuous therapy and structured treatment interruptions, *Chemical Engineering Science*, **64**, 1600–1617.
- [6] HUANG, Y.; WU, H. and ACOSTA, E.P. (2010). Hierarchical Bayesian inference for HIV dynamic differential equation models incorporating multiple treatment factors, *Biometrical Journal*, **52**, 470–486.
- [7] LIANG, H.; MIAO, H. and WU, H. (2010). Estimation of constant and time-varying dynamic parameters of HIV infection in a nonlinear differential equation model, *The Annals of Applied Statistics*, **4**, 460–483.
- [8] LUO, R., PIOVOSO, M.J., MARTINEZ-PICADO, J. and ZURAKOWSKI, R. (2012). HIV Model parameter estimates from interruption trial data including drug efficacy and reservoir dynamics, *PLoS ONE*, **7**, 7, e40198.
- [9] NOWAK, M.A. and MAY, R.M. (2000). *Virus Dynamics: Mathematical Principles of Immunology and Virology*, Oxford University Press, Oxford.
- [10] PERELSON, A.S. (2002). Modelling viral and immune system dynamics, *Nature Reviews Immunology*, **2**, 28–36.
- [11] PERELSON, A.S.; KIRSCHNER, D.E. and BOER, R. (1993). Dynamics of HIV infection of CD4⁺T cells, *Mathematical Biosciences*, **114**, 81–125.

- [12] RIVADENEIRA, P.S.; MOOG, C.H.; STAN, G.B.; COSTANZA, V.; BRUNET, C.; RAFFI, F.; FERRÉ, V.; MHAWAJ, M.J.; BIAFORE, F.; OUATTARA, D.A.; ERNST, D.; FONTENEAU, R. and XIA, X. (2014). Mathematical modeling of HIV dynamics after antiretroviral therapy initiation: a clinical research study, *AIDS Research and Human Retroviruses*, **30**, 831–834
- [13] ROCHA, D.; SCOTTO, M.; PINTO, C.; TAVARES, J. and GOUVEIA, S. (2019). Simulation study of HIV temporal patterns using Bayesian methodology, *Springer Proceedings in Mathematics & Statistics*, Springer (Ed.) (to appear).
- [14] STAFFORD, M.A., COREYA, L., CAO, Y., DAARDD, E.S., HOB, D.D. and PERELSON, A.S. (2000). Modeling plasma virus concentration during primary HIV infection, *Journal of Theoretical Biology*, **203**, 285–301.
- [15] WHITBY, L.; WHITBY, A.; FLETCHER, M.; HELBERT, M.; REILLY, J.T. and BARNETT, D. (2013). Comparison of methodological data measurement limits in CD4⁺T lymphocyte flow cytometric enumeration and their clinical impact on HIV management, *Cytometry Part B (Clinical Cytometry)*, **84**, 4, 248–254.