14th World Congress in Computational Mechanics (WCCM)
ECCOMAS Congress 2020
Virtual Congress: 11 – 15 January 2021
F. Chinesta, R. Abgrall, O. Allix and M. Kaliske (Eds)

APPROXIMATE BAYESIAN COMPUTATION APPLIED TO MODEL SELECTION AND PARAMETER CALIBRATION IN CELL PROLIFERATION

NILTON P. SILVA[1,2], BRUNA R. LOIOLA[3], JOSÉ M. J. COSTA[4] AND HELCIO R. B. ORLANDE[1,5]

[1] Federal University of Rio de Janeiro, UFRJ, Mechanical Engineering Program, COPPE
Cid. Universitária, Cx. Postal: 68503, Rio de Janeiro, RJ, 21941-972, Brazil

[2] Federal University of Amazonas, UFAM, Department of Mechanical Engineering
1200 General Rodrigo Otávio Avenue, Coroado I, Manaus, AM, 69080-900, Brazil
niltonps@ufam.edu.br

[3] Military Institute of Engineering, IME, Department of Mechanical Engineering
80 General Tibúrcio Square, Urca, Rio de Janeiro, RJ, 22290-270, Brazil
bruna.loiola@ime.eb.br

[4] Federal University of Amazonas, UFAM, Department of Statistics
1200 General Rodrigo Otávio Avenue, Coroado I, Manaus, AM, 69080-900, Brazil
zemir@ufam.edu.br

[5] Federal University of Rio de Janeiro, UFRJ, Oncobiology Program
Cid. Universitaria, Cx. Postal: 68503, Rio de Janeiro, RJ, 21941-972, Brazil
helcio@mecanica.coppe.ufrj.br

Key words: Inverse Problems, Approximate Bayesian Computation, cell proliferation.

Abstract. Approximate Bayesian Computation is used in this work for the selection and calibration of cell proliferation models. Four competing models based on ordinary differential equations are analyzed, by using the measurements of the proliferation of DU-145 prostate cancer viable cells during seven days. The selection criterion of the ABC algorithm is based on the Euclidean distance between the model prediction and the experimental observations. The Richards Model and the Generalized Logistic Model were selected by the ABC algorithm used in this work, providing accurate estimates of the evolution of the number of viable cells. Bayes factor revealed that there was no evidence in favor of any of these two selected models.

1 INTRODUCTION

Cell proliferation is numerically given by the difference between the numbers of newly-divided and dying cells. In order to predict the number of viable cells, several mathematical

models have been proposed in the literature [1,2]. These models have been applied for tumors, since in cancer cells the proliferation process is increased due to the abnormal metabolic activity [3]. Costa et al. [4], for example, have used one of these proliferation models to represent the behavior of prostate cancer cells (DU-145) *in vitro*. In addition, they have analyzed a chemotherapy treatment using doxorubicin (DOX). Costa et al. [5] have applied Approximate Bayesian Computation *via* a Monte Carlo Sequential Method (ABC-SMC) [6-8] to select from competing models the one that best represented the proliferation of prostate cancer tumor cells during *in vitro* chemotherapy experiments. Distinct hypotheses are included in a specific model. Thus, the selection and calibration of these models are of great interest.

The goal of this work is to select among four continuous models the one that better represents *in vitro* experimental data of the proliferation of DU-145 human prostate cancer cells. In order to perform this analysis, the Approximate Bayesian Computation (ABC) algorithm of Toni et al. [6] is applied for model selection and calibration, since this algorithm is robust and indicated for cases that the likelihood is not exactly known [4], such as in this work.

2 MATHEMATICAL MODELS

Different approaches can be used to model cell proliferation, by applying continuous, discrete or hybrid models. The choice of the model type for the investigation depends on the type of experiment, goal of the study and mainly the biological characteristics of the cells under analysis. In this work, the experimental data was obtained from ATCC [9], as shown in Figure 1.

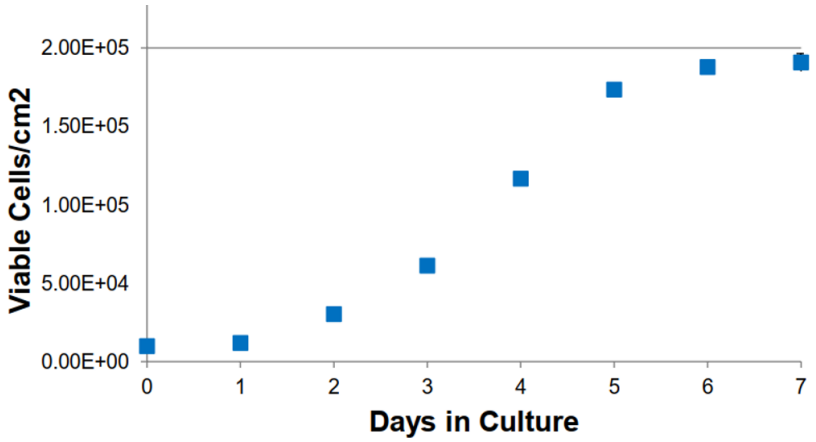

Figure 1: In vitro experiments results provided by American Type Culture Collection (ATCC) [9].

Due to the characteristics of the experimental data, without cycling or repeating behavior, only continuous models are used here in the inverse analysis of model selection/calibration. As can be seen in Figure 1, the number of viable cells did not grow without bounds. For this reason, the Exponential and Mendelsohn models were not considered in the analysis. In order to consider a bound in the proliferation process, four models are investigated: Logistic, Gompertz, Richards and Generalized Logistic [1-3]. The predictions provided by these models assume

uniform cell distribution and proliferation in the cell culture. In addition, the experiments are considered isothermal, at a constant temperature of 37 ˚C.

In these models, N_i is the number of viable cells varying with time t , with the initial number of cells given by $N_{0,i}$. The rate of proliferation is given by parameter α_i and the growth saturation by the dimensionless parameter γ_i. The support capacity that takes into account the space condition, oxygen availability and nutrient source is considered by $K_{sup,i}$. The mathematical models are presented in Equations 1-8 where the subscripts i = 1, 2, 3, 4 designate the models.

2.1 Model 1: Logistic Model

$$\frac{dN_1(t)}{dt} = \alpha_1 N_1(t)\left[1 - \frac{N_1(t)}{K_{sup_1}}\right]; \qquad t > 0 \tag{1}$$

$$N_1(0) = N_{0_1}; \qquad t = 0 \tag{2}$$

2.2 Model 2: Gompertz Model

$$\frac{dN_2(t)}{dt} = \alpha_2 N_2(t) \ln\left(\frac{K_{sup_2}}{N_2(t)}\right); \qquad t > 0 \tag{3}$$

$$N_2(0) = N_{0_2}; \qquad t = 0 \tag{4}$$

2.3 Model 3: Richards Model

$$\frac{dN_3(t)}{dt} = \alpha_3 N_3(t)\left[1 - \left(\frac{N_3(t)}{K_{sup_3}}\right)^{\gamma_3}\right]; \quad t > 0 \tag{5}$$

$$N_3(0) = N_{0_3}; \qquad t = 0 \tag{6}$$

2.4 Model 4: Generalized Logistic Model

$$\frac{dN_4(t)}{dt} = \frac{\alpha_4}{\gamma_4} N_4(t)\left[1 - \left(\frac{N_4(t)}{K_{sup_4}}\right)^{\gamma_4}\right]; \quad t > 0 \tag{7}$$

$$N_4(0) = N_{0_4}; \qquad t = 0 \tag{8}$$

3 SENSITIVITY ANALYSIS

Before the solution of the inverse problem, it is important to analyze the sensitivity coefficients of the measured variables with respect to each parameter. An analysis of the reduced sensitivity coefficients (X_r) is performed here, which are obtained by multiplying the parameter by the first partial derivative of the response with respect to that parameter [10-21]. The reduced sensitivity coefficients of viable cells with respect to the parameters are presented in Figure 2 for all the four models.

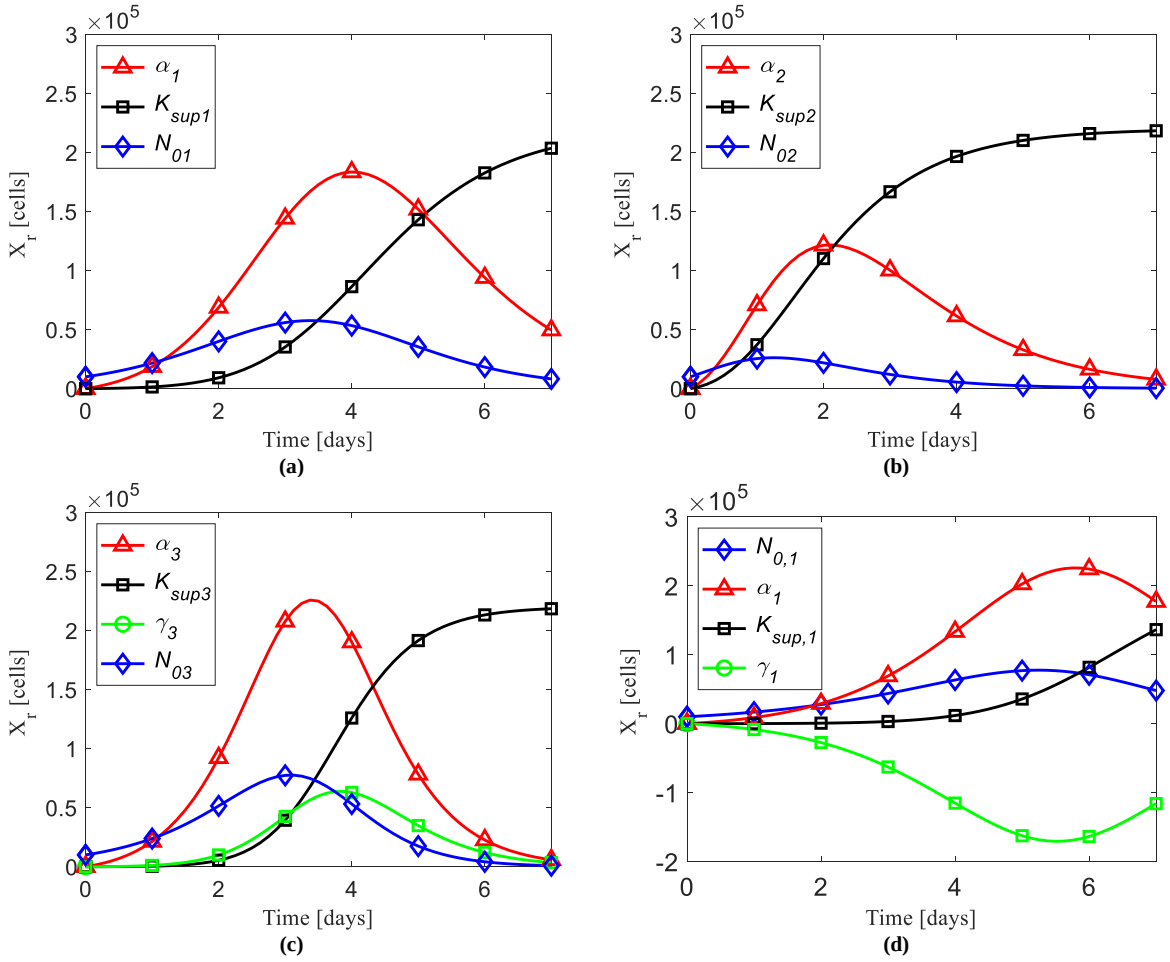

Figure 2: Reduced sensitivity coefficients: (a) Logistic Model, (b) Gompertz Model, (c) Richards Model, (d) Generalized Logistic Model

The reduced sensitivity coefficients were calculated with the parameter values given in Table 1. The sensitivity coefficients of the parameter N_0 for all models suddenly increase and then decay until approximately a null value, as time increases. The sensitivity coefficients with respect to the parameter K_{sup} increases until the steady state is reached. The sensitivity coefficients are not linearly dependent for the parameters of the Logistic model and of the Gompertz model. On the other hand, parameters α and γ of Richards model and Generalized Logistic model are correlated, as shown by Figures 2c and 2d.

Table 1: Nominal values used for the sensitivity analysis

Parameter	Value
N_0 [cell]	10,000
α [day^{-1}]	0.9
K_{sup} [cell]	220,000
γ	1.7

4 APPROXIMATE BAYESIAN COMPUTATION ALGORITHM

The Approximate Bayesian Computation (ABC) algorithm of Toni et al [6] was used in this work for the simultaneous model selection and estimation of the model parameters. This algorithm is presented in Table 2.

Table 2: ABC Algorithm [6]

1. Define the tolerances $\varepsilon_1, \varepsilon_2, \ldots \varepsilon_P$ for each of the iterations (populations) used for selecting the model and its parameters. Also, specify the distance function $d(\mathbf{Y},\mathbf{Y}^*)$ that substitutes the likelihood function. Set the population indicator $p = 0$.
2. Set the particle indicator $i = 1$, where each particle represents, at each iteration, a model and its parameters.
3. Sample the model M^* from the prior distribution for the models $\pi(\mathbf{M})$.
   If $p = 0$, sample the candidate parameters $\mathbf{P}^{**}$ from the prior distribution for the parameters of model M^*, that is, $\pi[\mathbf{P}(M^*)]$. Else, sample $\mathbf{P}^*$ from the parameters in the previous population $P(M^*)_{p-1}^i$, with weights $w(M^*)_{p-1}^i$, and perturb this particle to obtain $\mathbf{P}^{**} \approx K_p(\mathbf{P}^*, \mathbf{P}^{**})$, where K_p is a perturbation kernel.
4. If $\pi(\mathbf{P}^{**}) = 0$, return to step 3. Else, simulate from the forward problem (operator f) a candidate set of observable variables with model M^* and parameters $\mathbf{P}^{**}$, that is, $\mathbf{Y}^* = f(\mathbf{Y}|\mathbf{P}^{**}, M^*)$.
5. If $d(\mathbf{Y},\mathbf{Y}^*) > \varepsilon_p$, return to step 3. Otherwise, set $M_p^i = M^*$, add $\mathbf{P}^{**}$ to the population of particles $P(M^*)_p^i$ and calculate the particle weight
$$w(M^*)_p^i = \begin{cases} 1 & if\ p = 0 \\ \dfrac{\pi\left(P(M^*)_p^i\right)}{\sum_{j=1}^{N} w(M^*)_{p-1}^j K_p\left(P(M^*)_{p-1}^j, P(M^*)_p^i\right)} & if\ p > 0 \end{cases}$$
6. If $i < N$, where N is the number of particles, set $i = i + 1$ and go to step 3.
7. Normalize the weights.
8. If $p < P$, where P is the number of iterations (populations), set $p = p + 1$ and go to step 2. Otherwise, terminate the iterations.

Instead of using the likelihood function, the ABC algorithm is based on a distance function calculated at each set of successive populations (formed by particles composed of the model selected and parameters estimated). A tolerance (ε) is prescribed at each population for the distance function given in this work by the Euclidean distance between the system dependent variable $\mathbf{Y}^*$ and the experimental data $\mathbf{Y}$. If the Euclidean distance is smaller than the tolerance,

the particle is accepted; otherwise, the particle is rejected and a new particle is generated.

5 RESULTS AND DISCUSSIONS

The ABC-SMC method with 4000 particles was applied to model selection and estimation of the model parameters. Uniform priors and uniform transition Kernels were adopted for the parameters, as presented in Table 3. Note that the upper and lower limits of the transition Kernel were assumed to be ± 1% of the upper bound of the prior for each parameter.

Table 3: ABC Priors distribution and transition Kernels for the parameters

Models	Priori	Transition Kernel
1,2,3,4	$N_0 \sim U(1{,}000;19{,}000)$	U(-190;190)
1,2,3,4	$\alpha \sim U(0.09;4.5)$	U(-0.045;0.045)
1,2,3,4	$K_{sup} \sim U(22{,}000;418{,}000)$	U(-4180;4180)
3,4	$\gamma \sim U(0.170;8.5)$	U(-0.085;0.085)

The experimental data presented in Figure 1, obtained from ATCC™ (American Type Culture Collection) [9] for the number of viable cells of DU-145 prostate cancer cells during *in-vitro* experiments, were used in the inverse problem. In order to solve the inverse problem, the four mathematical models presented in Equations 1 to 8 were solved by the 4th order Runge-Kutta algorithm. The choice of the tolerances for the sequential populations of the ABC algorithm were set by starting at 5.42×10^5 and finishing at 5.42×10^3, along a total of fifty-seven populations. The last tolerance was imposed in accordance with Morozov's discrepancy principle, assuming a standard deviation (σ) of 1% of the maximum value of viable cells, that is, $\varepsilon_{last} = \sigma\sqrt{N_{measurements}}$.

The ABC-SMC algorithm of Toni et al. [6] selected the Richards Model and the Generalized Logistic Model, as shown by Figure 3, to represent the experimental data presented in figure 1. In this figure it is possible to observe that after 27 populations only these two models have been selected. However, a total of 57 populations were needed to perform the correct calibration of the parameters of these two models.

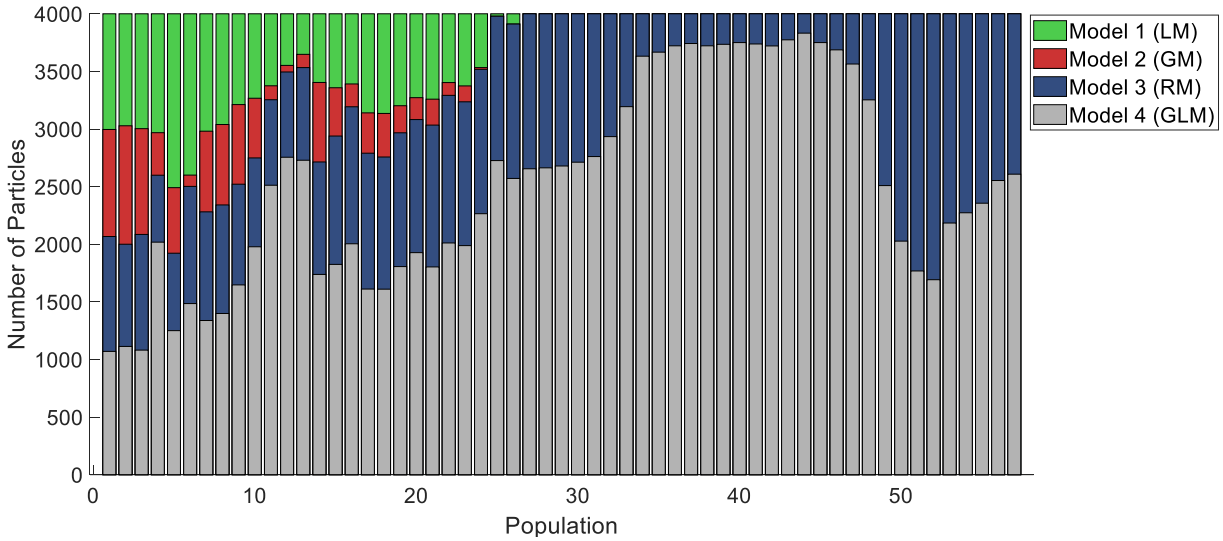

Figure 3: Number of accepted particles at each population

The problem of selecting models can be associated to hypothesis tests, such as the Bayes factor proposed by Kass and Raftery [22]. The Bayes Factor for Models 3 and 4 is given by the posterior probability of each model in relation to the data, that is,

$$B_{43} = \frac{\pi(M_4|\mathbf{Y})}{\pi(M_3|\mathbf{Y})} \tag{9}$$

The criteria of Kass and Raftery [22] for interpreting the Bayes factor is presented in Table 4. At the final population shown in Figure 3, 2608 particles were selected for Model 4 and 1392 were selected for Model 3, which gives a Bayes factor of 1.87. In accordance with Table 4, there is no evidence in favor of any of the models 3 or 4.

Table 4: Bayes factor [22]

B_{43}	$2\ln(B_{43})$	Evidence against Model 3
1 to 3	0 to 2	Not worth more than a bare mention
3 to 20	2 to 6	Positive
20 to 150	6 to 10	Strong
>150	> 10	Very strong

The histograms of the model parameters at the final population are presented by Figures 4 and 5. These histograms exhibit approximate Gaussian behaviors, centered at mean values. The means, standard deviations and 95% credible intervals for the estimated parameters for both models 3 and 4 are presented in Table 5 and 6, respectively.

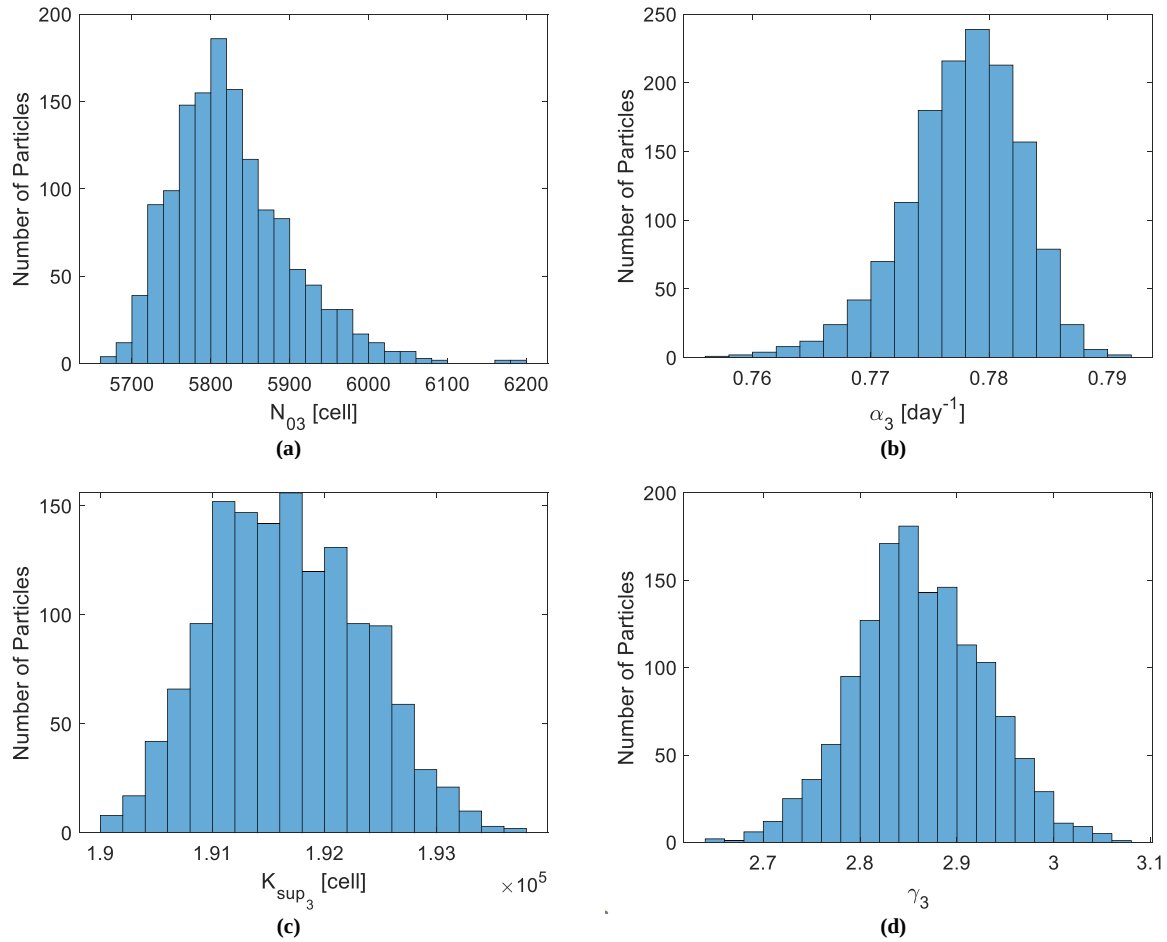

Figure 4: Histograms for the parameters of Model 3: (a) N_{03}, (b) α_3, (c) K_{sup3}, (d) γ_3

The numbers of viable cells computed with models 3 and 4, considering the mean of the accepted particles at the last population, are presented in figures 6 and 7, respectively. The light blue lines in these figures are the estimated curves calculated with each of the accepted particles at the final population. These figures show that both model estimations have an excellent agreement with the experimental data, thus confirming that either one of the competing models 3 or 4 (Richards or Generalized Logistic) could be used to represent in vitro experiments performed with DU-145 human prostate cancer cells.

We note that the results presented here were not influenced by the stochastic simulations performed. In fact, the results were qualitatively unchanged in four runs of the ABC-SMC algorithm used in this work.

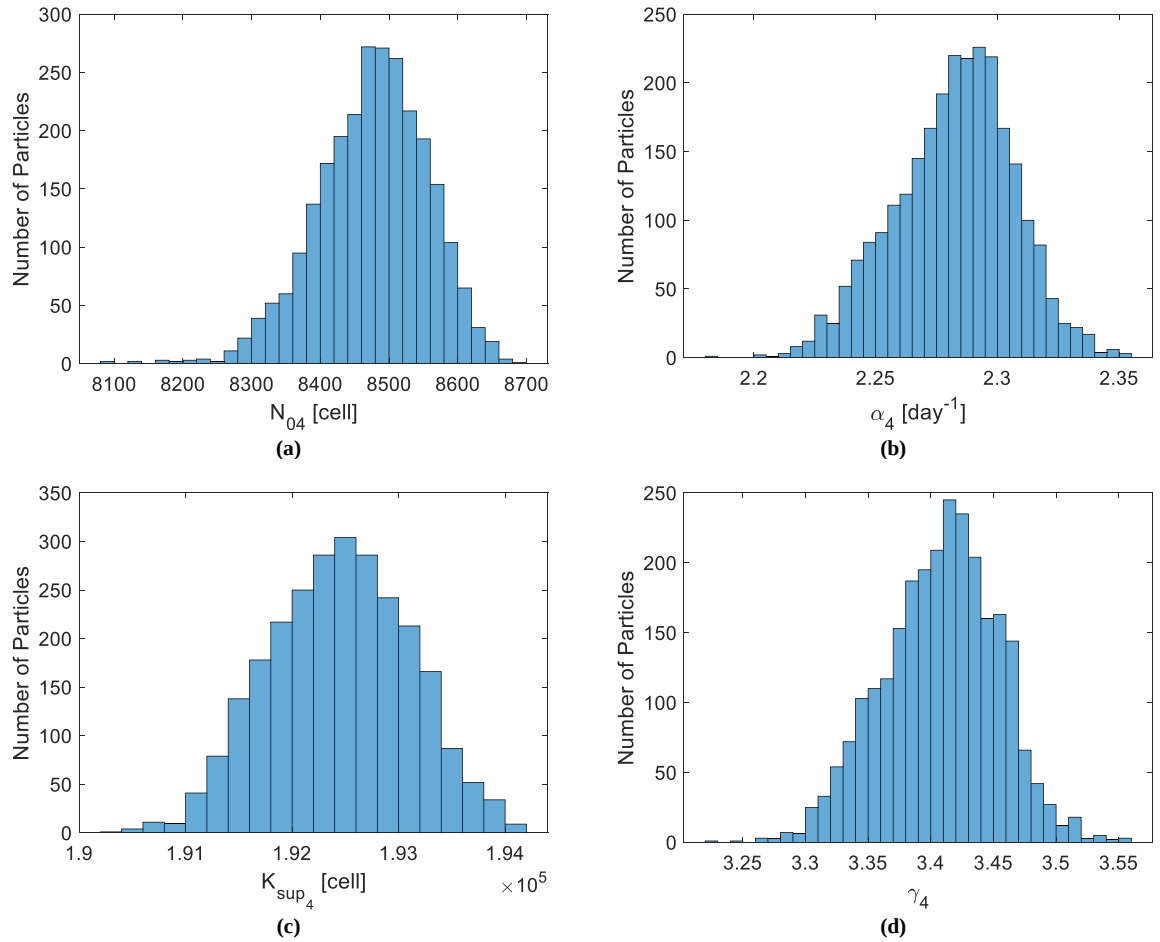

Figure 5: Histograms for the parameters of Model 4: (a) N_{04}, (b) α_4, (c) K_{sup4}, (d) γ_4

Table 5: Parameters estimation - Model 3

Parameter	Mean	Standard deviation	Lower limit 95%	Upper limit 95%
N_{03} [cell]	5826.6	74.9072	5712.7	5998.6
α_3 [day^{-1}]	0.7777	0.0049	0.7667	0.7858
K_{sup3} [cell]	191,660	671.9028	190,460	193,040
γ_3	2.8621	0.0662	2.7315	2.9938

Table 6: Parameters estimation - Model 4

Parameter	Mean	Standard deviation	Lower limit 95%	Upper limit 95%
N_{04} [cell]	8476.1	80.0335	8309.2	8615.0
α_4 [day^{-1}]	2.2823	0.0244	2.2316	2.3273
K_{sup4} [cell]	192,440	656.2398	191,200	193,700
γ_4	3.4078	0.0457	3.3164	3.4912

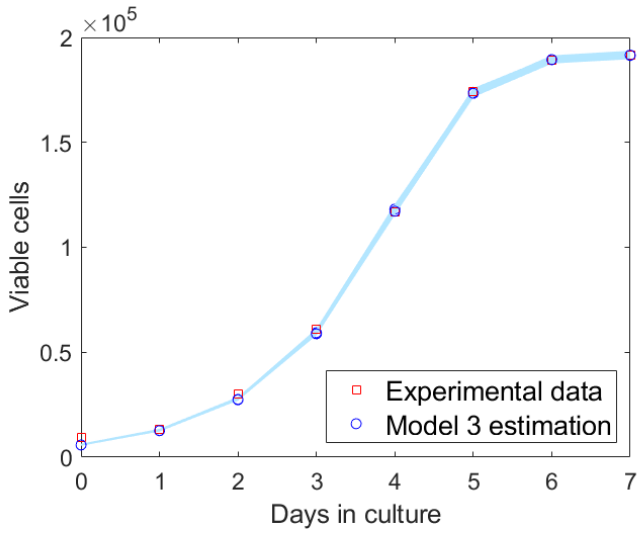

Figure 6: Comparison between experimental data and Model 3 estimation

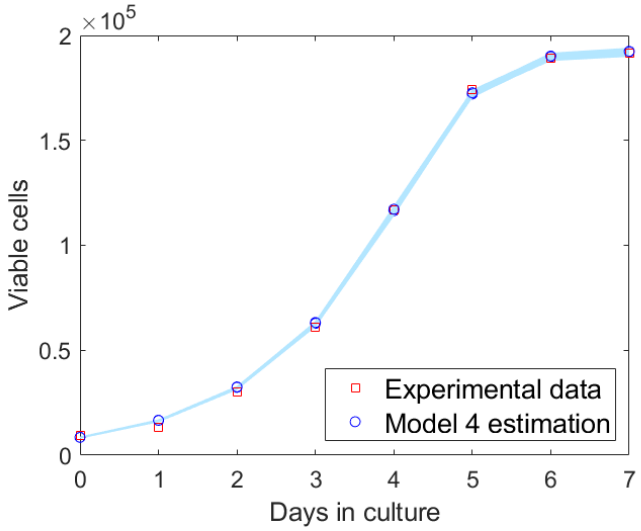

Figure 7: Comparison between experimental data and Model 4 estimation

6 CONCLUSIONS

The ABC-SMC algorithm of Toni et al. [6] was applied with 4000 particles for model selection and estimation of cell proliferation model parameters. The parameters were considered with uniform priors and uniform transition Kernels were used in the algorithm. In order to solve the inverse problem, the four mathematical models were solved with the Runge-Kutta´s 4th order method. The tolerances for the sequential populations of ABC-SMC method

decreased from $5.42x10^5$ to $5.42x10^3$ along fifty-seven populations. The last tolerance was imposed in accordance with the assumed measurement uncertainty following Morozov's discrepancy principle. The Richards Model and the Generalized Logistic Model were both selected by ABC-SMC algorithm, providing accurate estimations of the number of viable cells. An analysis of Bayes factor revealed that both models can be used to accurately represent *in vitro* measurements of the time evolution of the DU-145 human prostate cancer cells.

ACKNOWLEDGEMENTS

Authors acknowledge CNPq, CAPES, FAPEAM and FAPERJ for the financial support of this work.

REFERENCES

[1] Preziosi, L., *Cancer Modelling and Simulation*. Boca Raton: CRC Press LLC, (2003).

[2] Murphy, H., Jaafari, H. and Dobrovolny, H.M. Differences in predictions of ODE models of tumor growth: a cautionary example. BMC Cancer (2016) 16:163

[3] Otto, S.P. and Day, T. *A Biologist's Guide to Mathematical Modeling in Ecology and Evolution*. Princeton University Press, (2011).

[4] da Costa, J.M.J., Orlande H.R.B., and da Silva W.B.. Model selection and parameter estimation in tumor growth models using Approximate Bayesian Computation-ABC. *Computational and Applied Mathematics* (2018) 37(3):2795-2815.

[5] da Costa, J.M.J., Orlande, H.R.B. Lione, V.O.F. Lima, A.G.F. Cardoso T.C.S. and Varón, L.A.B. Simultaneous Model Selection and Model Calibration for the Proliferation of Tumor and Normal Cells During In Vitro Chemotherapy Experiments. *Journal of Computational Biology* (2018) 25:1-16.

[6] Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P.H. Approximate Bayesian Computation scheme for parameter inference and model selection in dynamical systems. *Journal of The Royal Society Interface* (2009) 6(31):187-202.

[7] Liepe, J., Kirk, P., Filippi, S., Toni, T., Barnes, C.P. and Stumpf, M.P.H.. A framework for parameter estimation and model selection from experimental data in systems biology using approximate Bayesian computation. *Nature protocols* (2014) 9:439-456.

[8] Toni T. and Stumpf, M.P.H. Simulation-based model selection for dynamical systems in systems and population biology. *Bioinformatics* (2010) 26(1):104-110.

[9] ATCC: Thawing, Propagating, and Cryopreserving Protocol, NCI-PBCF-HTB81 (DU 145) Prostate Carcinoma (ATCC®HTB-81™) February 27, (2012). Version 1.6.

[10] Alifanov, O.M. *Inverse Heat Transfer Problems*, Springer, Berlin Heidelberg, (1994).

[11] Alifanov, O.M. Artyukhin, E.A. and Rumyantsev, S.V. *Extreme Methods for Solving Ill-Posed Problems with Applications to Inverse Heat Transfer Problems*, Begell House, New York, 1995.

[12] Dulikravich, G.S. Shape inverse design and optimization for tree-dimensional aerodynamics, AIAA invited paper 95–0695, AIAA *Aerospace Sciences Meeting*, Reno, NV, USA, January (1995).

[13] Kurpisz, K. and Nowak, A.J. *Inverse Thermal Problems*, WIT Press, Southampton, UK,

(1995).
[14] Dulikravich, G.S. Design and optimization tools development, in: H. Sobieczky (Ed.), *New Design Concepts for High Speed Air Transport*, Springer Wien/New York, (1997), pp. 159–236.
[15] Trujillo, D.M. and Busby, H.R. *Practical Inverse Analysis in Engineering*. CRC Press, Boca Raton, FL, (1997).
[16] Bertero, M. and Boccacci P., *Introduction to Inverse Problems in Imaging*. Institute of Physics, London, UK, (1998).
[17] Denisov, A.M. *Elements of the Theory of Inverse Problems*. VSP, Utrecht, The Netherlands, (1999).
[18] Yagola, A.G., Kochikov, I.V., Kuramshina, G.M. and Pentin, Y.A. *Inverse Problems of Vibrational Spectroscopy*. VSP, Utrecht, The Netherlands, (1999).
[19] Ozisik, M.N and Orlande, H.R.B. *Inverse Heat Transfer: Fundamentals and Applications*. Taylor and Francis, New York, (2000).
[20] Woodbury K.A., *Inverse Engineering Handbook*. CRC Press, Boca Raton, (2002).
[21] Kaipio, J. and Somersalo, E. *Statistical and Computational Inverse Problems*. Springer-New York, (2004).
[22] Kass, R. E. and Raftery, A. E. Bayes Factors. *Journal of the American Statistical Association* (1995) 90(430):773-795.