24th COBEM - 2017

24th ABCM International Congress of Mechanical Engineering
December 3-8, 2017, Curitiba, PR, Brazil

COBEM-2017-0814
BAYESIAN APPROACH TO PARAMETER ESTIMATION AND SELECTION OF GROWTH MODELS OF TUMOR CELLS

Adriana Machado Malafaia da Mata
Federal University of Espírito Santo, Vitória, ES
drika.m3@hotmail.com
José Mir Justino da Costa
Federal University of Amazonas, Manaus, AM
zemir@ufam.edu.br
Wellington Betencurte da Silva and Julio Cesar Sampaio Dutra
Federal University of Espírito Santo, Campus de Alegre, Alegre, ES
wellingtonufes@gmail.com, juliosdutra@yahoo.com.br

Abstract. *Cancer is a disease that arises from the disordered growth of cells. Commonly, antineoplastic chemotherapy is used to treat the most common cancers. In this context, researches have turned to mathematical models that describe the growth of tumor cells with an action of a chemotherapeutic drug. Faced with a variety of models, a method for selecting the most suitable model has become promising. This paper studies mathematical models of tumor cell growth and applies the Approximate Bayesian Computation (ABC) method based on the Sequential Monte Carlo (SMC) to select the best model that fits the observed data. A two-compartment pharmacokinetic model allowed the study of orally administered antineoplastic drugs. In addition, ABC SMC was able to estimate the parameters of the selected model.*

Keywords: *Approximate Bayesian Computation (ABC), Sequential Monte Carlo (SMC), models of tumor cell growth, Bayesian model selection.*

1. INTRODUCTION

According to the World cancer report 2014 of the International Agency for Research on Cancer of the World Health Organization, it is expected that the impact of cancer on the population of developing countries will correspond to 80 % of more than 20 million of new cases estimated for 2025. From the National Cancer Institute, Brazil estimates about 600.000 new cases of cancer occur in 2016-2017.

From the 1950, chemotherapy has become one of the leading weapons of cancer medicine (Teixeira and Fonseca, 2007). On Rodrigues (2011), anti-neoplastic chemotherapy requires greater quantitative and analytical understanding. For this study, modeling is presented as an alternative for a better understanding of the optimal dosage of the chemotherapeutic drug. Since the correct dosage should control the growth of tumor cells and affect the minimum possible of normal cells in the patient's body.

Many models have been proposed to describe the growth of tumor cells. Malthus (1798) presented a population growth model in which the growth rate is proportional to the population at a given moment. This model is called Exponential and describes well the early stages of a tumor. Another model successfully applied to a large number of biological phenomena is the Logistic model, which includes the limit for cell growth represented by the carrying capacity K (Gerlee, 2013). Richards (1959) proposed a flexible growth function for empirical use to deal with situations in which the growth curve is asymmetric. Mendelsohn (1962) introduced a general equation. Laird (1964) proposed the application of the Gompertz curve (growth rate model of mortality) in the study of tumor growth. Von Bertalanffy (1968) proposed the development of the organism in a systemic way. Herman *et al.* (2011) integrated the relationships among growth of solid tumors, metabolic rate, vascularization and necrosis.

Models describing the chemotherapeutic treatment of a tumor include to the tumor growth equation a term of effect of the drug in the treatment. Rodrigues (2011) focused on the investigation of antiangiogenic protocols. Pinho *et al.* (2013) considered interactions between normal cells, cancer cells, endothelial cells and antiangiogenic agent. The Modified Exponential model (da Costa, 2015) was proposed for a problem of dynamic modeling of tumor growth in vitro. However, these models only consider single-compartment approach covering drugs administered by infusion. Two-compartment pharmacokinetic modeling allowed the study of antineoplastic drugs administered orally (Gallo Neto, 2012).

Generally, for biological systems, reliable information about the parameters is lacking (Toni *et al.*, 2009). In addition, there are a variety of models that describe cell growth. In this regard, this present works concerns about the choice of the

model that best represents the measured data using ABC SMC algorithm. The results shows that this method has correctly selected the mathematical models.

2. METHODOLOGY

Pharmacokinetics of two-compartment models proposed by Gallo Neto *et al.* (2012) shown in equations 1 and 2, are used here, where the mass balances in the central (1) and plasma (2) compartments.

$$\frac{dm_1}{dt} = qC(t) - K_a m_1 \tag{1}$$

$$\frac{dm_2}{dt} = K_a m_1 - K_{et} m_2 \tag{2}$$

Where K_a and K_{et} refer to the absorption and elimination constants respectively, m_1(mg) is the mass of the drug present in the central compartment and m_2(mg) is the mass of the drug in the plasma compartment. The amount of drug administered as a function of time is defined by $qC(t)$. The subscripts $i = 1,2,3$ then refer to the model.
As da Costa *et al.* (2017), this study was carried out using the Generalized Logistic model, the Gompertz model and the Modified Exponential model.
The pharmacokinetics of the three models used were bicompartmental, so Equations 1 and 2 integrate the mathematical modeling of all models used.

2.1 Generalized Logistic model

Based on Richards (1959) which considered an additional parameter (γ) for tumor growth formulation, Equation 3 refers to the chemotherapy treatment of a tumor by the Generalized Logistic model.

$$\frac{dN_1}{dt} = \frac{\alpha_1}{\gamma} N_1 \left(1 - \frac{N_1}{K_1}\right)^{\gamma} - \frac{\mu_1 N_1 m_2}{a_1 + N_1} \tag{3}$$

The index 1 corresponds to Generalized Logistic model; $N(t)$ = number of cells at time t; α = rate of cell growth; K = cell support capacity, ie the maximum number of cells that a tumor can reach with available nutrients and limited space; γ = parameter that makes it possible to adjust cell growth asymmetrically. When $\gamma = 1$ growth occurs logistically, symmetrically; μ = treatment rate of cells; a = constant of Holling type 2 (Pinho *et al.*, 2013).

2.2 Gompertz model

From the Gompertz curve proposed by Laird (1964) in tumor growth, Equation 4 refers to the chemotherapeutic treatment of a tumor by the Gompertz model.

$$\frac{dN_2}{dt} = \alpha_2 N_2 ln\left(\frac{K_2}{N_2}\right) - \frac{\mu_2 N_2 m_2}{a_2 + N_2} \tag{4}$$

The index 2 corresponds to Gompertz model; $N(t)$ = number of cells at time t; α = rate of cell growth; K = cell support capacity, ie the maximum number of cells that a tumor can reach with available nutrients and limited space; μ = treatment rate of cells; a = constant of Holling type 2 (Pinho *et al.*, 2013).

2.3 Modified Exponential model

In the early stages of tumor growth the cells grow exponentially. The Modified Exponential model was proposed by da Costa (2015) on a problem of dynamic modeling of tumor growth in vitro. In this Modified Exponential model, even when the initial condition is $N_i(0) = 0$ there will be an increase in tumor cells, which is impossible. Therefore, a direct bootstrap method was applied consistently with Differential Algebraic Equations systems based on Vieira and Biscaia (2001). A regularization parameter was chosen so that there was no cell growth when $N_i(0) = 0$, this is presented in Equation 5. The exponential model was then used according to Equation 6.

The parameters $a_1 = 10^{-3}$ and $a_2 = 10^{-5}$ values were adjusted so that the hyperbolic tangent equals one (+1) and $n = 1$ when the initial condition $N_i(0) \neq 0$; the hyperbolic tangent will be (-1) and $n = 0$ when $N_i(0) = 0$.

$$n = \frac{\left(1 + tanh\left(\frac{N_i(0) - a_1}{a_2}\right)\right)}{2} \tag{5}$$

$$\frac{dN_3}{dt} = n\left(\alpha_3 K_3 e^{-\alpha_3 t} - \frac{\mu_3 N_3 m_2}{a_3 + N_3}\right) \tag{6}$$

The index 3 corresponds to Modified Exponential model; $N(t)$ = number of cells at time t; α = rate of cell growth; K = cell support capacity, ie the maximum number of cells that a tumor can reach with available nutrients and limited space; μ = treatment rate of cells; a = constant of Holling type 2 (Pinho *et al.*, 2013).

Each model is subject to the following initial conditions:

$$N_i(0) = N_{i0}, i = 1,2,3 \tag{7}$$

$$m_1(0) = m_{10} \tag{8}$$

$$m_2(0) = m_{20} \tag{9}$$

ABC methods are applied to infer posterior distributions without the need for likelihood functions, which are replaced by a comparison between observed and simulated data (Toni *et al.*, 2009). Specifically, for each sampled model, parameter estimation are performed using a ABC SMC algorithm for model selection. The model with the highest probability will be the largest number of particles accepted in the final population, guaranteeing a good estimation of the posterior distribution for the parameters. The algorithm of the ABC SMC method is detailed in Box 1, where we can observe that the tolerance establishes the limit for the Euclidean distance between observed and simulated data. The initial tolerance should be high and reduced to each population. The particle is chosen randomly among the candidate particles of all models. A uniform kernel is used to disturb the particle. The particle is accepted when the probability of the disturbed particle is nonzero and the Euclidean distance between observed and simulated data is less than the tolerance. It is established a number of accepted particles that is desired for each population.

Box 1. ABC SMC algorithm

1. Initialize tolerances $\varepsilon_1, \varepsilon_2, ..., \varepsilon_p$.
   Set the indicator population p=0.
2. Define the indicator particle $i = 1$.
3. Sample M^* from $\pi(M)$.
   If $p = 0$ sample θ^{**} independently from $\pi(\theta(M^*))$.
   If $p > 0$ sample θ^* of the priori population $\{\theta(M^*)_{p-1}\}$ with weight $\omega(M^*)_{p-1}$.
   Perturb θ^* to get $\theta^{**} \sim K_p(\theta|\theta^*)$.
   If $\pi(\theta^{**}) = 0$, return to 3.
   Simulate a set of candidate data $Z^* \sim \pi(Z|\theta^{**}, M^*)$.
   If $d(Z^*, Z_0) \geq \varepsilon_p$, return to 3.
4. Define $M_p^{(i)} = M^*$ and add θ^{**} for the population of particles $\{\theta(M^*)_p\}$ and calculate the weights as $\omega_p^{(i)} = 1$, if $p = 0$ and

   $$\omega_p^{(i)} = \frac{\pi(\theta^{**})}{\sum_{j=1}^{N} \omega_{p-1}^{(j)} K_p(\theta_{p-1}^{(j)}|\theta^{**})}, \text{ if } p > 0.$$

5. If $i < N$ define $i = i + 1$ and return to 3.
6. For each model M, normalize the weights of the accepted particles.
7. If $p < P$ define $p = p + 1$ and return to 2.

3. RESULTS AND DISCUSSIONS

Measurements with noise were generated for each model using Matlab. The nominal values for the parameters are shown in Table 1. The ABC SMC algorithm was implemented with uniform distribution. Each model is subject to initial conditions given by $N_{i0} = 10^3; m_1 = 5.4352mg; m_2 = 0$.

For analysis of the ABC algorithm with experimental data of tumor cells with treatment, as measurements generated from the Generalized Logistic model (model 1), were used, as a result of the direct problem and a random noise with a standard deviation of 5 % was added. To use the ABC algorithm *a priori* used for the parameters was generated from uniform distributions from literature data, according to Table 2.

Tolerance sets the limit for the Euclidean distance between observed and simulated data. To select a model that best fits the measured data, initial tolerance should be high to avoid high rejection rates and reduced to each population. The most likely model *a posteriori* will be the one with the highest number of particles accepted in the final population, ensuring a good estimate of the posterior distribution for the parameters.

Table 1: Parameters that generated measurements.da Costa *et al.* (2017).

Parameter	Nominal value
α_i	2.3×10^{-1}
K_i	2×10^{5}
μ_i	1.1×10^{2}
a_i	7.5×10^{2}
Ka_i	6×10^{-2}
Ket_i	5×10^{-2}

Table 2: *A priori* distribution of the parameters.

Parameter	Distribution
α_i	$U(0,1)$
K_i	$U(10^3, 5 \times 10^5)$
γ	$U(1,2)$
μ_i	$U(1, 2 \times 10^2)$
a_i	$U(5 \times 10^2, 10^3)$
Ka_i	$U(10^{-2}, 10^{-1})$
Ket_i	$U(10^{-3}, 10^{-1})$

In this test, the initial tolerance was 10^{13} and in the last population (10), the tolerance was 10^{11}, in which only particles of the Generalized Logistic model were accepted, indicating the selection of model 1. It can be seen from Figure 1 that ABC correctly selected the model that generated the measurements. However, for this case it is observed that the model was selected faster than in the other tests, with a tolerance still very close to the initial one. But it is observed that the obtained results, when compared to the measured, do not present the same agreement as in the other cases. Possibly more intermediate populations are needed to obtain better estimates.

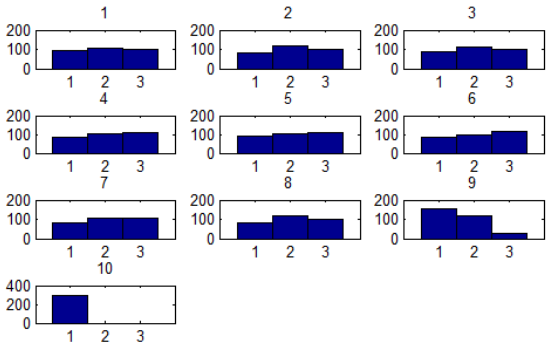
The results presented in Table 3 show a good estimate using the ABC algorithm, comparing the average of each parameter. However, particle dispersion is observed in relation to most of the parameters. In this case, the selection of the model occurred with a tolerance still very high, generating dispersion among the particles of the final population.

Table 3: Estimation of the parameters for the data generated from the Generalized Logistic model of Tumor Cells - with treatment.

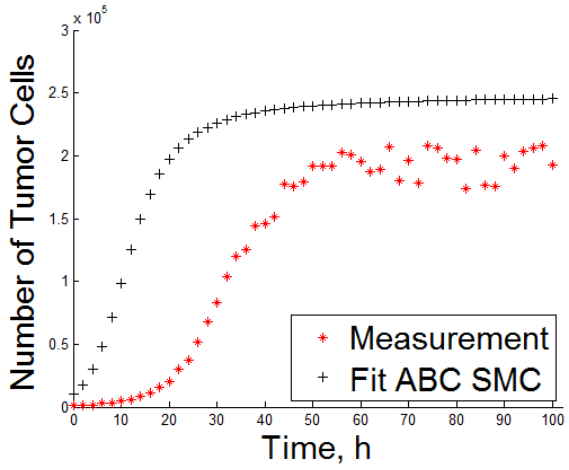
Population	Parameter	Quantil 0.01	Mean	Quantil 0.99
10	α_1	2.067×10^{-1}	3.915×10^{-1}	6.835×10^{-1}
	K_1	1.461×10^{5}	2.035×10^{5}	2.901×10^{5}
	γ	1.220	1.609	1.891
	μ_1	2.689×10^{1}	9.779×10^{1}	1.745×10^{2}
	a_1	5.825×10^{2}	7.543×10^{2}	9.004×10^{2}
	Ka_1	2.597×10^{-2}	5.488×10^{-2}	8.511×10^{-2}
	Ket_1	1.546×10^{-2}	4.914×10^{-2}	8.647×10^{-2}

Also in the model selection analysis, measurements were generated from the Gompertz model (model 2) with a noise, with a standard deviation of 5 %. For the use of the ABC algorithm *a priori* used for the parameters, uniform distributions were chosen in accordance with the Table. 2.
The initial tolerance was 10^{13} and in the last population (31), the tolerance was 6.6×10^{9}, in which only particles of the Gompertz model were accepted, confirming the correct selection of the model that generated the measurements (model 2), as seen in the Figure 2.

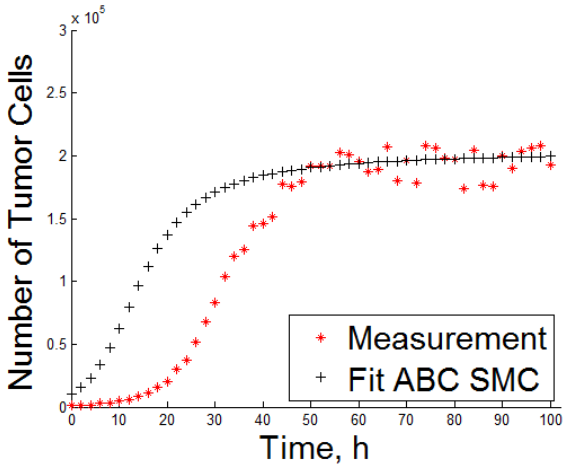
The results presented in Table 4 show a good agreement between the values measured and estimated through the ABC algorithm, in relation to the average of each parameter. However, the parameter μ shows greater dispersion between the results of the accepted particles in the final population.



(a) Selected particles in each population.



(b) First population.



(c) Last population (10).

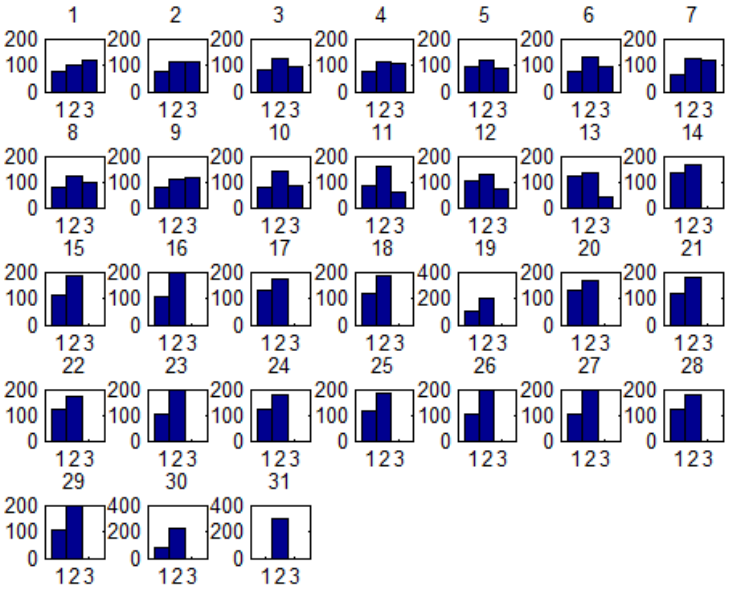
Figure 1: Generalized Logistic (model 1) tumor cells with chemotherapy.

Table 4: Estimation of the parameters for the data generated from the Gompertz model of Tumor Cells - with treatment.

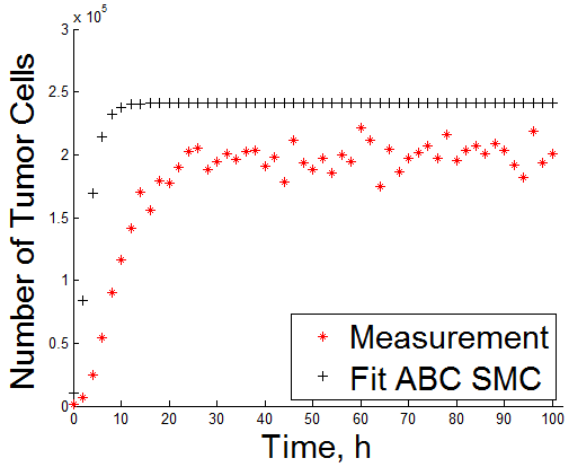
Population	Parameter	Quantil 0.01	Mean	Quantil 0.99
31	α_2	2.235×10^{-1}	2.387×10^{-1}	2.549×10^{-1}
	K_2	1.988×10^{5}	2.021×10^{5}	2.057×10^{5}
	μ_2	5.822×10^{1}	1.098×10^{2}	1.595×10^{2}
	a_2	6.220×10^{2}	7.535×10^{2}	8.823×10^{2}
	Ka_2	3.447×10^{-2}	5.595×10^{-2}	7.933×10^{-2}
	Ket_2	2.371×10^{-2}	4.917×10^{-2}	7.513×10^{-2}

Finally the evaluation of the model selection, the measurements were generated from the Modified Exponential model (model 3) with addition of a random noise with standard deviation of 5 %. The *a priori* candidate particles were generated by uniform distributions according to Table 2.
The initial tolerance was 10^{13} and the final tolerance was 6×10^{9}, when all the accepted particles belong to the Modified Exponential model, indicating the selection of model 3. It can be seen from Figure 3 that ABC correctly selected the model that generated the measurements. It is also observed that as the tolerances are being refined the estimated values are gradually approaching the measurements.

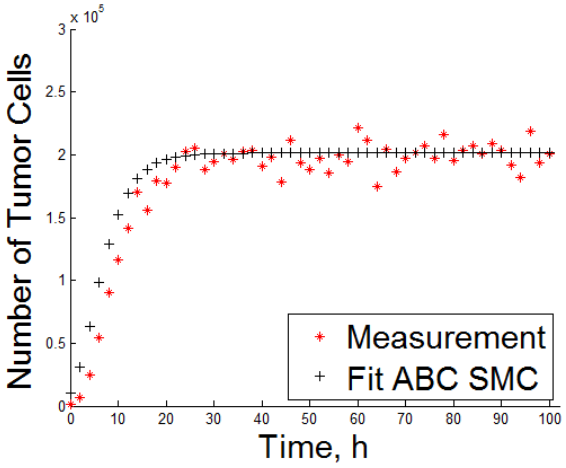
The results presented in Table 5 show a good accuracy between the measured values and the average of the parameters estimated by means of the algorithm ABC in the last population. The parameter μ (treatment rate) shows a greater



(a) Selected particles in each population.



(b) First population.



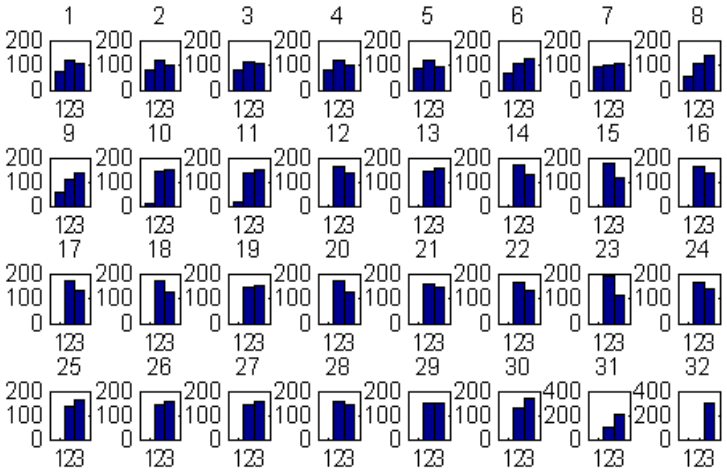
(c) Last population (31).

Figure 2: Gompertz (model 2) of tumor cells with chemotherapy.

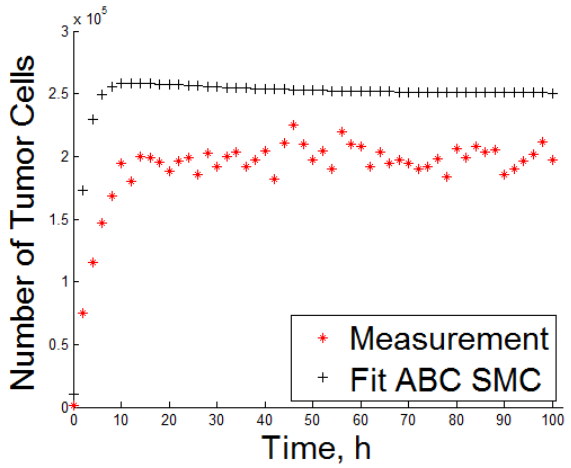
dispersion among the quantile results, in relation to the other parameters.

Table 5: Estimation of the parameters for the data generated from the Modified Exponential model of Tumor Cells - with treatment.

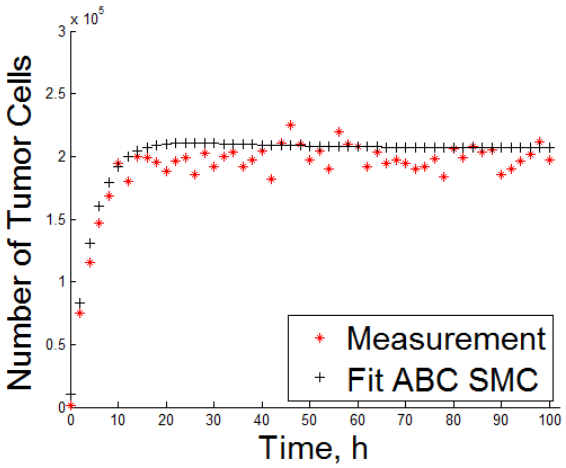
Population	Parameter	Quantil 0.01	Mean	Quantil 0.99
32	α_3	1.656×10^{-1}	2.207×10^{-1}	2.900×10^{-1}
	K_3	2.001×10^{5}	2.057×10^{5}	2.121×10^{5}
	μ_3	5.946×10^{1}	9.582×10^{1}	1.361×10^{2}
	a_3	6.231×10^{2}	7.539×10^{2}	8.960×10^{2}
	Ka_3	3.831×10^{-2}	6.022×10^{-2}	8.234×10^{-2}
	Ket_3	3.569×10^{-2}	5.956×10^{-2}	8.093×10^{-2}



(a) Selected particles in each population.



(b) First population.



(c) Last population (32).

Figure 3: Modified Exponential (model 3) tumor cells with chemotherapy.

4. CONCLUSION

ABC SMC was applied for selecting the model that originated the simulated data. This method selected correctly the mathematical model of tumor growth in Two-compartment pharmacokinetic modeling. This verification allows the application of ABC SMC to select models of tumor development, with the influence of chemotherapy and extends the studies for drugs administered orally.

This study allowed to verify the need to adjust the tolerances in each case, to avoid high rates of particle rejection.

5. REFERENCES

da Costa, J.M.J., 2015. *Problema de estimativa de estado e de estimativa simultânea de modelos e parâmetros em crescimento de tumores*. Ph.D. thesis, Universidade Federal do Rio de Janeiro.

da Costa, J.M.J., Orlande, H.R.B. and da Silva, W.B., 2017. "Model selection and parameter estimation in tumor growth models using approximate bayesian computation-abc".

Gallo Neto, M., 2012. *Modelagem farmacocinética e análise de sistemas lineares para a predição da concentração de medicamentos no corpo humano*. Ph.D. thesis, Universidade de São Paulo.

Gallo Neto, M., Ferreira, M.S. and Yanagihara, J.I., 2012. "Development and validation of a pharmacokinetic model

using linear system analysis for the prediction of medicaments distribution in the human body".

Gerlee, P., 2013. "The model muddle: in search of tumor growth laws". *Cancer research*, Vol. 73, No. 8, pp. 2407–2411.

Herman, A.B., Savage, V.M. and West, G.B., 2011. "A quantitative theory of solid tumor growth, metabolic rate and vascularization". *PloS one*, Vol. 6, No. 9, p. e22973.

Laird, A.K., 1964. "Dynamics of tumour growth". *British journal of cancer*, Vol. 18, No. 3, p. 490.

Mendelsohn, M.L., 1962. "Autoradiographic analysis of cell proliferation in spontaneous breast cancer of c3h mouse. iii. the growth fraction 2". *Journal of the National Cancer Institute*, Vol. 28, No. 5, pp. 1015–1029.

Pinho, S.T.R.d., Bacelar, F.S., Andrade, R.F.S. and Freedman, H., 2013. "A mathematical model for the effect of anti-angiogenic therapy in the treatment of cancer tumours by chemotherapy". *Nonlinear Analysis: Real World Applications*, Vol. 14, No. 1, pp. 815–828.

Richards, F., 1959. "A flexible growth function for empirical use". *Journal of experimental Botany*, Vol. 10, No. 2, pp. 290–301.

Rodrigues, D.S., 2011. "Modelagem matemática em câncer: dinâmica angiogênica e quimioterapia anti-neoplásica".

Teixeira, L.A. and Fonseca, C.O., 2007. *De doença desconhecida a problema de saúde pública: o INCA e o controle do câncer no Brasil*. Ministério da Saúde, Rio de Janeiro, Brasil.

Toni, T., Welch, D., Strelkowa, N., Ipsen, A. and Stumpf, M.P., 2009. "Approximate bayesian computation scheme for parameter inference and model selection in dynamical systems". *Journal of the Royal Society Interface*, Vol. 6, No. 31, pp. 187–202.

Vieira, R. and Biscaia, E., 2001. "Direct methods for consistent initialization of dae systems". *Computers & Chemical Engineering*, Vol. 25, No. 9, pp. 1299–1311.

Von Bertalanffy, L., 1968. "General system theory". *New York*, Vol. 41973, No. 1968, p. 40.

6. RESPONSIBILITY NOTICE

The authors are the only responsible for the printed material included in this paper.