Modeling breast tumor growth by a randomized logistic model: A computational approach to treat uncertainties via probability densities

Clara Burgos-Simón[1]*, Juan-Carlos Cortés[1],
David Martínez-Rodríguez[1], Rafael-J. Villanueva[1]

[1] Instituto Universitario de Matemática Multidisciplinar,
Universitat Politècnica de València, Valencia, Spain

email: clabursi@upv.es, jccortes@upv.es,
damarro3@.upv.es, rjvillan@upv.es

Abstract

We consider a randomized discrete logistic equation to describe the dynamics of breast tumor volume. We propose a method, that takes advantage of the principle of maximum entropy, to assign reliable distributions to model inputs (initial condition and coefficients) and sample data, respectively. Since the distributions of coefficients depend on certain parameters, we design a computational procedure to determine the above mentioned parameters using the information of the probabilistic distributions. The proposed method is successfully applied to model the breast tumor volume using real data. The approach seems to be flexible enough to be adapted to other stochastic models in future contributions.

Keywords: Maximum entropy principle; Computational model fitting; Volume tumor growth; Uncertainty treatment.

1 Introduction

Breast cancer is one of the most common malignant diseases in the female population, around 1/8 of women are affected by this illness. It is the second most commonly diagnosed cancer in women worldwide, [1, 2]. Over the last decades, this type of cancer is increasing due to several reasons: the enlargement of the life expectancy and, consequently, the increase of DNA mutations, and the current unhealthy lifestyle (physical inactivity, obesity, living in polluted areas, etc.). The breast cancer is the first cause of death by malignant tumors in the female population aged between 40 and 59. Nevertheless, in the recent years, the survival of this malignancy has been increased because of new therapies and the early prevention and prediction [3].

A key point in the early prevention of breast cancer is the capacity of measure the volume of tumors and predict their growth over the time. To quantify the volume of tumors, doctors use approximate measurement techniques based on medical images via electronic devices, [4, 5]. These measurements involve intrinsic errors in the real volume dimension that must be taken into account. As it shall be seen later, errors can be modelled by applying the principle of maximum entropy (PME), that allows us to allocate reliable uncertainties to sampled tumor volume data [6].

*Corresponding author

To study and predict the growth of volume tumors cancer several nonlinear mathematical models, based on difference and differential equations, have been successfully proposed. In [7], the authors develop a nonlinear system of difference equations to study the short term dynamics of the bladder cancer and the immune response of patients. In [8] a numerical scheme for solving time-fractional cancer invasion system with non-local diffusion has been recently proposed. In this paper, authors propose an optimal control strategy to enhance the power of NK-cells and Efector T-cells in order to more quickly eradicate the cancer. In [9], authors perform a numerical analysis to understand the dynamics of cancer invasion using a time-fractional system. In [10], the classical Gomperzian growth is applied to study the breast tumor volume before applying suitable therapies. In [11], the logistic model is parametrized to predict the treatment response and changes in breast cancer cellularity during neoadjuvant chemotherapy. Cancer developement is a process where normal somatic cells acquire mutations, in [12], a system of nonlinear differential equations to study the dynamics of these cells mutations is proposed. Recently, the dynamics of a cell line (MCF-7) in human breast cancer has been described using the same type of mathematical formulation, [2].

As it has been previously pointed out, tumor volume data involve uncertainties, then it is natural to consider stochastic models to study the evolution over the time of the breast tumors volume. In this contribution we consider a randomized logistic-type model to study the growth of breast tumors volume. Despite of its simple formulation, this class of models have demonstrated to be very effective to describe the dynamics of biological growth processes like tumors [3].

In order to incorporate uncertainties to the logistic model, one usually distinguish two main approaches, namely, stochastic differential equations (SDEs) and random differential equations (RDEs). SDEs are usually forced by processes such as a Wiener process or Brownian motion whose sample behavior is highly irregular (non-differentiable sample paths). The rigorous treatment of SDEs requires special stochastic calculus like Itô or Malliavin [13, 14]. Complementary, RDEs are those in which random effects are directly manifested in model parameters (initial/boundary conditions, forcing or source term and/or coefficients) that are assumed random variables or stochastic processes with regular behavior (e.g., sample continuous or differentiable) with respect to time and/or space [15]. RDEs have demonstrated to be flexible models to quantify uncertainty in mathematical modelling since a wide variety of probability distributions can be allocated for each model parameter rather than using a global stochastic process, like the Wiener process, to include uncertainties in the whole model. In this paper, we will consider a logistic-type RDE whose initial condition and coefficients are random variables whose probability distributions must be consistently set from sampled information.

Indeed, in dealing with practical applications of RDEs, as modelling breast tumors volume using real data, a main challenge is allocating appropriate probability distributions for model parameters so that the output model, which is a stochastic process, satisfactorily captures data uncertainties. In this paper we face this key challenge by developing a computational technique to quantify uncertainties and then performing more realistic predictions to modelling breast tumors volume by means of a random logistic equation using real data. Assuming randomness to each model parameter (initial condition and coefficients), this computational technique is based on seeking the best probability density distribution (PDF) of model parameters so that the PDF of the solution stochastic process of the random logistic model matches the PDF assigned, via PME, to sampled data of breast tumor volume at each time instant. In this manner, through the PDF, we perform a more complete probabilistic description of the breast tumor volume dynamics than constructing predictions based only on the expectation and confidence intervals.

At this point, it is important to emphasize that when applying RDEs to modeling real problems, the allocation of appropriate probability distributions to model parameters is often done using heuristic arguments based on positiveness, boundedness and/or meta-information [16, 17]. This limits the choice of distributions to particular families. For instance, for positive and bounded parameters, the Beta distribution may be an appropriate candidate; for positive and unbounded parameters, the Gamma distribution might be suitable; etc. In contrast, the PME method allows us to give more flexibility when assigning probability distributions to each model parameters, since a parametric family of distributions are seeking for.

Our analysis will be presented in the following steps. Section 2 is devoted to introduce two auxiliary results. In Subsection 2.1, the randomized discrete logistic model is presented together with the expression of the PDF of its solution stochastic process. In Subsection 2.2, the PME is described as a suitable method to assign a PDF when only limited sampling information is available. In Section 3 we will apply the PME to assign an explicit PDF to each sampled data. In Section 4 we will again utilize the PME to represent the PDF of each model input via closed expressions, which depend on certain parameters to be determined later. In Section 5 we design a computational procedure to determine the aforementioned parameters so that the density of the model solution be as close as possible to the density previously allocated to sampled data. In Section 6 we will apply the computational approach, introduced in the previous section, to first obtain the densities of model inputs and, secondly of the model output. Finally, conclusions are drawn in Section 7.

2 Auxiliary stochastic results

This section is addressed to introduce some technical results, about the randomized logistic model and the PME, that will be required through the paper.

2.1 A randomized discrete logistic model

The logistic model has been extensively applied to describe the dynamics of growth processes in different scientific areas, as pharmacology [18], epidemiology [19] or ecology [20]. In this paper we are interested in its application in medicine to model tumor growth [11, 21].

In this contribution we consider the following discrete dynamical system, usually referred to as the Pielou model [22, 23],

$$X_{n+1} = \frac{AX_n}{1+BX_n}, \quad n = 0, 1, 2, \dots, \tag{1}$$

for a given initial condition X_0. As it can be seen in [22], this discrete model comes from the classical Verhulst continuous logistic equation [24]

$$V'(t) = aV(t)\left(1 - \frac{V(t)}{b}\right), \tag{2}$$

being $a > 0$ the growth rate and $b > 0$ the carrying capacity. According to [22, pp. 19–22], models (1) and (2) are related via the following relationship of their respective parameters,

$$A = e^a, \qquad B = \frac{e^a - 1}{b}. \tag{3}$$

Since $a > 0$ and $b > 0$, then $A > 1$ and $B > 0$.

As it has been pointed out in the foregoing section, uncertainty quantification is a main goal in modeling breast tumors growth from real data. This aim us at treating the parameters A, B and X_0 in model (1) as random variables belonging to a complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$. As a consequence, model parameters depend on outcomes $\omega \in \Omega$, i.e., $A = A(\omega)$, $B = B(\omega)$ and $X_0 = X_0(\omega)$, and then the solution is a stochastic process, $X_n = X_n(\omega)$. As usual, hereinafter the ω-notation will be hidden.

Although properties of random quantities are often described via statistical moments like the mean and the variance, it is more desirable to do it through probability distributions. Specifically, fixed t, from the so called first PDF, $f_Y(y,t) := f_Y(y)$, of a stochastic process, say $Y(t)$, one can calculate the mean, $\mathbb{E}[Y(t)]$, the variance, $\mathbb{V}[Y(t)] = \mathbb{E}[(Y(t))^2] - (\mathbb{E}[Y(t)])^2$, and also any higher one-dimensional statistical moment of arbitrary order $m = 1, 2, \ldots$, [25, Ch. 3]

$$\mathbb{E}[Y(t)^m] = \int_{-\infty}^{\infty} y^m f_Y(y) \mathrm{d}y,$$

as well as to construct confidence intervals and also to calculate the probability that the process lies in a specific interval of interest

$$\mathbb{P}[y_1 \leq Y(t) \leq y_2] = \int_{y_1}^{y_2} f_Y(y) \mathrm{d}y.$$

By applying the random transformation technique [25], recently some of the authors have obtained an explicit expression of the first PDF, f_{X_n}, to the solution of the randomized Pielou model (1), [26]. Specifically, by assuming that A, B and X_0 are absolutely continuous random variables with a joint PDF, $f_{X_0,A,B}$, they obtained

$$f_{X_n}(x) = \int_{\mathcal{D}(A,B)} f_{X_0,A,B}\left(\frac{x(a-1)}{a^n(a-1) - bx(a^n-1)}, a, b\right) \left|\frac{(a-1)^2 a^n}{(a^n(a-1) - bx(a^n-1))^2}\right| \mathrm{d}a\mathrm{d}b, \tag{4}$$

where $\mathcal{D}(A,B)$ denotes the domain of random vector (A,B), [26]. In the particular case that A, B and X_0 are independent, then $f_{X_0,A,B}(x_0,a,b) = f_{X_0}(x_0) f_A(a) f_B(b)$ (being f_{X_0}, f_A and f_B the PDF of X_0, A and B, respectively) and, as a consequence, the PDF of the solution can be represented as

$$f_{X_n}(x) = \int_{\mathcal{D}(A,B)} f_{X_0}\left(\frac{x(a-1)}{a^n(a-1) - bx(a^n-1)}\right) f_A(a) f_B(b) \left(\frac{(a-1)^2 a^n}{(a^n(a-1) - bx(a^n-1))^2}\right) \mathrm{d}a\mathrm{d}b. \tag{5}$$

Computing this double integral in an exact way, i.e. using primitives, is not always possible. Nevertheless, using numerical quadrature rules we can approximate it. This fact mainly depends upon the mathematical expression of the densities f_{X_0}, f_A and f_B. To overcome this drawback, we will consider the following representation of f_{X_n} in terms of the expectation operator, $\mathbb{E}[\cdot]$,

$$f_{X_n}(x) = \mathbb{E}\left[f_{X_0}\left(\frac{x(A-1)}{A^n(A-1) - Bx(A^n-1)}\right) \left|\frac{(A-1)^2 A^n}{(A^n(A-1) - Bx(A^n-1))^2}\right|\right]. \tag{6}$$

At this point, it is important to underline that we can weak the condition that input parameters A and B are absolutely continuous random variables but just having probability distributions. Then, the density f_{X_n} can be computed using Monte Carlo simulations. However, notice that to follow this strategy, we need to assign reliable distributions to random variables A, B and X_0. This key point will be addressed using the PME and it will be described in the next subsection.

2.2 Principle of Maximum Entropy (PME)

This section is devoted to briefly describe and adapt the PME to our modeling problem. The mathematical concept of entropy is a measurement of uncertainty. It defines the lack of knowledge of a random variable, which has been built on the basis of limited probabilistic information. The larger the uncertainty of a random variable the larger its entropy. Specifically, the PME that we will use in this paper utilizes the concept of

Shannon's Entropy, S_Y, as a measure of uncertainty. This measure is defined as

$$S_Y = -\int_{\mathcal{D}(Y)} f_Y(y) \log(f_Y(y))\mathrm{d}y, \tag{7}$$

where $\mathcal{D}(Y)$ and f_Y denote, respectively, the domain and the PDF of the absolutely continuous random variable Y. Seeking for the function f_Y that maximizes S_Y can be interpreted as finding out the PDF of Y that corresponds to the maximal randomness and the minimal quantity of information. The latter, is usually given via the statistical moments (mean, variance, symmetry, kurtosis, etc.), the support, etc., [6, Chapter 2.2].

In our setting, the PME will be applied to assign reliable densities both to the sample data and the random input parameters A, B and X_0.

According to PME, the density f_Y is obtained by maximizing the functional S_Y subject to the available probabilistic information about the random variable, usually trough the statistical moments m_k, $k = 1, \ldots, K$, as well as imposing that the integral of the density f_Y on its domain, $\mathcal{D}(Y)$, is the unit

$$\int_{\mathcal{D}(Y)} f_Y(y)\mathrm{d}y = 1, \quad \mathbb{E}[Y^k] = \int_{\mathcal{D}(Y)} y^k f_Y(y)\mathrm{d}y = m_k, \quad k = 1, \ldots, K. \tag{8}$$

Using variational calculus, it can be seen that f_Y takes the following exponential form

$$f_Y(y) = \mathbb{1}_{\mathcal{D}(Y)} \mathrm{e}^{-1-\sum_{k=0}^{K} \lambda_k y^k}, \tag{9}$$

where $\mathbb{1}_{\mathcal{D}(Y)}$ denotes the characteristic function of $\mathcal{D}(Y)$, and the parameters λ_k, are determined solving the nonlinear system (8) once the moments m_k, $k = 1, 2, \ldots, K$, have been determined usually from a sample.

In our modelling setting, we will apply the PME method using sample information of the mean (m_1) and the variance ($m_2 + m_1^2$), hence $K = 2$. This will be done in Sections 3 and 4.

3 Data and their uncertainty

As it has been previously indicated, in this section we will apply the PME to assign probability distributions to each sampled data. To this end, we are going to use the following information. First, the figures tabulated in the second column of Table 1, that correspond to the sampled data of the breast tumor volume measured in mm^3, at different days, $\tilde{n}$, [27, Figure 1]. They have been obtained using xenograft technique, which consists of inserting cell tissue from one species to another, in our case, breast tumoral tissue from human species to a rodent species, [27, p. 2]. These values have been collected by measurement electronic devices, hence involving uncertainties. This fact aims us at treating these quantities as random variables rather than deterministic values. The figures $\tilde{m}_{1,\tilde{n}}$ are taken as representing the mean and, according to [28], we assign a variance of 5% at each value, i.e. $\tilde{\sigma}_{\tilde{n}}^2 = 0.05\, \tilde{m}_{1,\tilde{n}}$ (see third column, $\tilde{\sigma}_{\tilde{n}}^2$, Table 1). As a consequence, the second moment can be straightforwardly computed, $\tilde{m}_{2,\tilde{n}} = \tilde{m}_{1,\tilde{n}}^2 + \tilde{\sigma}_{\tilde{n}}^2$, see last column of Table 1.

Since we have information of the two first moments, we allocate the PDF of each the corresponding volume of breast tumor cell using expression (9) with $K = 2$, i.e.

$$\tilde{f}_{\tilde{n}}(x) = \mathbb{1}_{\mathcal{D}(\tilde{n})} \mathrm{e}^{-1-\lambda_0^{\tilde{n}}-\lambda_1^{\tilde{n}} x-\lambda_2^{\tilde{n}} x^2}, \tag{10}$$

where $\mathcal{D}(\tilde{n})$ denotes the domain of the random variable inferred by the information collected in Table 1 and $\lambda_k^{\tilde{n}}$, $k = 0, 1, 2$, are determined solving the following system of nonlinear equations for each $\tilde{n} \in$

Days	Mean ($\tilde{m}_{1,\tilde{n}}$)	Variance ($\tilde{\sigma}^2_{\tilde{n}}$)	2nd moment ($\tilde{m}_{2,\tilde{n}}$)
$\tilde{n} = 0$	45.74	2.287	2094.4
$\tilde{n} = 16$	194.257	9.7129	37745.49
$\tilde{n} = 30$	675.14	38.2570	455852.27
$\tilde{n} = 33$	936.53	46.8256	877135.26
$\tilde{n} = 43$	1941.7	97.0850	3770295.97
$\tilde{n} = 48$	2558.6	127.930	6546561.89

Table 1: Volume of breast tumor cells using xenograft technique at different days, [27] ($\tilde{m}_{1,\tilde{n}}$) together with the assigned variance ($\tilde{\sigma}^2_{\tilde{n}}$) and second moment ($\tilde{m}_{2,\tilde{n}}$).

$\{0, 16, 30, 33, 43, 48\}$,

$$\begin{aligned}
&\int_0^\infty \mathrm{e}^{-1-\lambda_0^{\tilde{n}}-\lambda_1^{\tilde{n}}x-\lambda_2^{\tilde{n}}x^2}\,\mathrm{d}x = 1,\\
&\int_0^\infty x\mathrm{e}^{-1-\lambda_0^{\tilde{n}}-\lambda_1^{\tilde{n}}x-\lambda_2^{\tilde{n}}x^2}\,\mathrm{d}x = \tilde{\mu}_{\tilde{n}},\\
&\int_0^\infty x^2\mathrm{e}^{-1-\lambda_0^{\tilde{n}}-\lambda_1^{\tilde{n}}x-\lambda_2^{\tilde{n}}x^2}\,\mathrm{d}x = \tilde{\mu}^2_{\tilde{n}} + \tilde{\sigma}^2_{\tilde{n}}.
\end{aligned} \tag{11}$$

The results are shown in Table 2. They have obtained by *fsolve* function in MATLAB, [29].

Days	$\lambda_0^{\tilde{n}}$	$\lambda_1^{\tilde{n}}$	$\lambda_2^{\tilde{n}}$
$\tilde{n} = 0$	74.5841	-3.2063	3.50e-02
$\tilde{n} = 16$	21.7070	-0.1882	4.8407e-04
$\tilde{n} = 30$	10.5921	-0.0133	8.6701e-06
$\tilde{n} = 33$	11.8918	-0.0145	8.0943e-06
$\tilde{n} = 43$	9.5029	-0.0034	9.3958e-07
$\tilde{n} = 48$	8.8842	-0.0015	3.1900e-07

Table 2: Values of $\lambda_0^{\tilde{n}}, \lambda_1^{\tilde{n}}$ and $\lambda_2^{\tilde{n}}$ obtained solving the system of nonlinear equations given in (11) the different values of $\tilde{n}$.

In Fig 1, we show a graphical representation of each PDF given by equation (10) with the values collected in Table 2. We can observe the PDFs built via the PME provide higher variability as $\hat{n}$ increases in full agreement with the variance $\tilde{\sigma}^2_{\tilde{n}}$ given in Table 1.

4 Statistical distribution of the model parameters

Once probability distributions to sampled data have been assigned, as it has been indicated in the Introduction section, the following step will consist of establishing probability distributions for model parameters A, B and X_0. To achieve this goal, the PME will be applied again.

For consistency with the distributions assigned in Section 3 for the first sampled data, corresponding to $\tilde{n} = 0$, we take

$$f_{X_0}(x_0) = \mathrm{e}^{-1-\lambda_0^0-\lambda_1^0 x_0-\lambda_2^0 x_0^2}, \tag{12}$$

where $\lambda_0^2 = 74.5841$, $\lambda_1^0 = -3.2063$ and $\lambda_2^0 = 3.50e-02$, see first row in Table 2.

Using PME, we propose the following parametric PDFs for the rest of random variables A and B

$$f_A(a) = \mathrm{e}^{-1-\lambda_0^A-\lambda_1^A a-\lambda_2^A a^2}, \quad a \in [a_1, a_2], \tag{13}$$

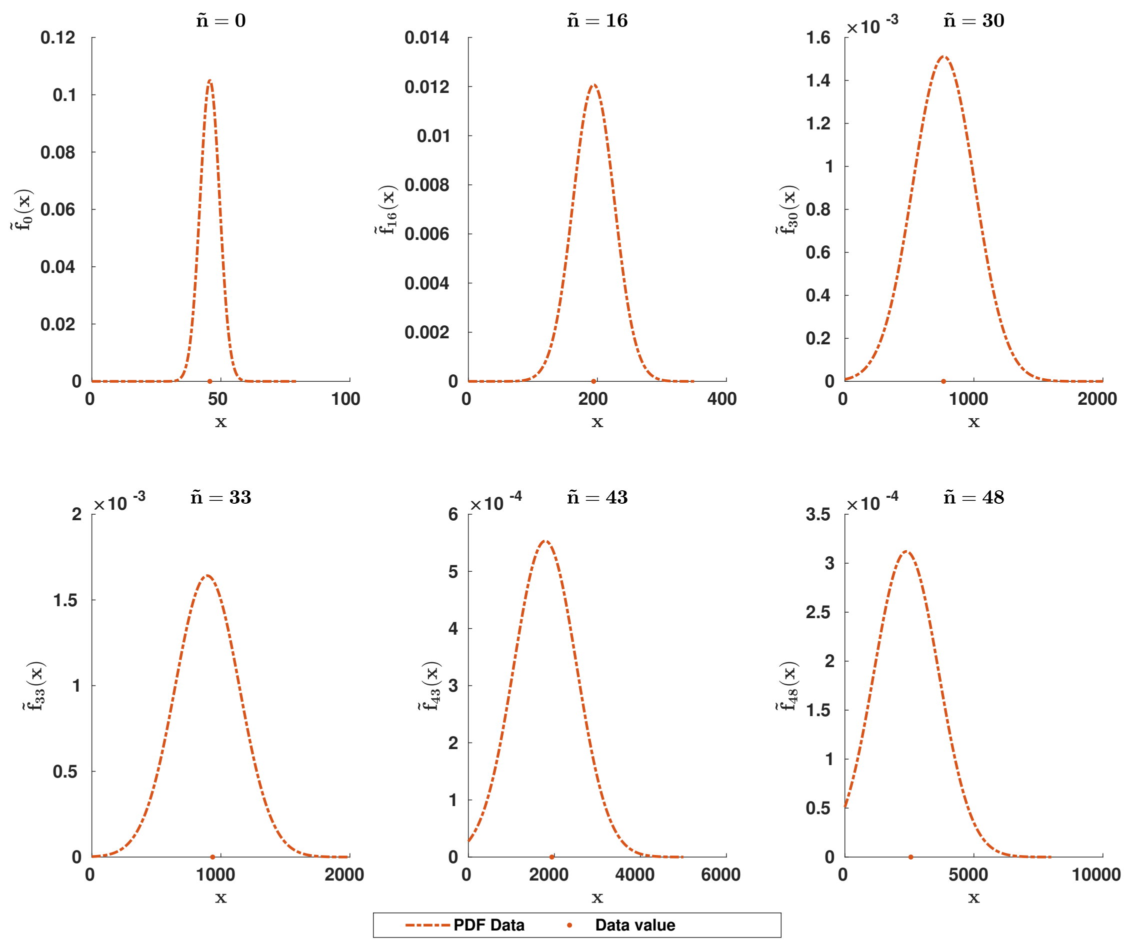

Figure 1: PDF of each sampled data using the PME at the days $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$. The red points represent the values $\tilde{m}_{1,\tilde{n}}$ given in Table 1.

$$f_B(b) = \mathrm{e}^{-1-\lambda_0^B-\lambda_1^B b-\lambda_2^B b^2}, \quad b \in [b_1, b_2], \tag{14}$$

respectively. According to (3), we derive that $a_1 = 1$ and $b_1 = 0$, hence $a_2 > 1$ and $b_2 > 0$.

The values of parameters $\left\{\lambda_0^A, \lambda_1^A, \lambda_2^A\right\}$ and $\left\{\lambda_0^B, \lambda_1^B, \lambda_2^B\right\}$ must be chosen so that f_A and f_B integrate the unit. Therefore, after calculating the integral and isolating λ_0^A and λ_0^B one gets, respectively,

$$\lambda_0^A = -1 + \frac{(\lambda_1^A)^2}{4\lambda_2^A} + \log\left[\frac{\sqrt{\pi}}{2\sqrt{\lambda_2^A}}\left(\mathrm{Erf}\left(\frac{\lambda_1^A + 2a_2\lambda_2^A}{2\sqrt{\lambda_2^A}}\right) - \mathrm{Erf}\left(\frac{\lambda_1^A + 2a_1\lambda_2^A}{2\sqrt{\lambda_2^A}}\right)\right)\right], \tag{15}$$

$$\lambda_0^B = -1 + \frac{(\lambda_1^B)^2}{4\lambda_2^B} + \log\left[\frac{\sqrt{\pi}}{2\sqrt{\lambda_2^B}}\left(\mathrm{Erf}\left(\frac{\lambda_1^B + 2b_2\lambda_2^B}{2\sqrt{\lambda_2^B}}\right) - \mathrm{Erf}\left(\frac{\lambda_1^B + 2b_1\lambda_2^B}{2\sqrt{\lambda_2^B}}\right)\right)\right], \tag{16}$$

provided $\lambda_2^A > 0$ and $\lambda_2^B > 0$. Here, $\mathrm{Erf}(x) = \frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}\mathrm{d}t$ stands for the error function.

According to expression (6), to compute the PDF of the solution stochastic process X_n, it is necessary to sampling random variables A and B. This will be done via the inverse of the distribution functions of A and B using the so called inverse transformation method, [30, Chapter 2]. According to this technique, we first need to calculate the distribution functions of A and B,

$$F_A(a) = \int_1^a f_A(s)\mathrm{d}s = \frac{1}{2\sqrt{\lambda_2^A}}\mathrm{e}^{-1-\lambda_0^A+\frac{(\lambda_1^A)^2}{4\lambda_2^A}}\sqrt{\pi}\left(-\mathrm{Erf}\left(\frac{\lambda_1^A + 2\lambda_2^A}{2\sqrt{\lambda_2^A}}\right) + \mathrm{Erf}\left(\frac{\lambda_1^A + 2a\lambda_2^A}{2\sqrt{\lambda_2^A}}\right)\right) \tag{17}$$

and

$$F_B(b) = \int_0^b f_B(s)\mathrm{d}s = \frac{1}{2\sqrt{\lambda_2^B}}\mathrm{e}^{-1-\lambda_0^B+\frac{(\lambda_1^B)^2}{4\lambda_2^B}}\sqrt{\pi}\left(-\mathrm{Erf}\left(\frac{\lambda_1^B}{2\sqrt{\lambda_2^B}}\right) + \mathrm{Erf}\left(\frac{\lambda_1^B + 2b\lambda_2^B}{2\sqrt{\lambda_2^B}}\right)\right), \tag{18}$$

where $1 \le a \le a_2$ and $0 \le b \le b_2$, respectively. Denoting $u_A := F_A(a) \in (0,1)$ and $u_B := F_B(b) \in (0,1)$ in (17) and (18), respectively, and isolating a and b in each expression, one gets

$$a = \frac{1}{2\lambda_2^A}\left(-\lambda_1^A + 2\sqrt{\lambda_2^A}\,\mathrm{InvErf}\left(\frac{2\mathrm{e}^{1+\lambda_0^A-\frac{(\lambda_1^A)^2}{4\lambda_2^A}}\sqrt{\pi}\,u_A\sqrt{\lambda_2^A} + \pi\mathrm{Erf}\left(\frac{\lambda_1^A+2\lambda_2^A}{2\sqrt{\lambda_2^A}}\right)}{\pi}\right)\right), \tag{19}$$

and

$$b = \frac{1}{2\lambda_2^B}\left(-\lambda_1^B + 2\sqrt{\lambda_2^B}\,\mathrm{InvErf}\left(\frac{2\mathrm{e}^{1+\lambda_0^B-\frac{(\lambda_1^B)^2}{4\lambda_2^B}}\sqrt{\pi}\,u_B\sqrt{\lambda_2^B} + \pi\mathrm{Erf}\left(\frac{\lambda_1^B}{2\sqrt{\lambda_2^B}}\right)}{\pi}\right)\right), \tag{20}$$

respectively. Here $\mathrm{InvErf}(\cdot)$ denotes the inverse function of $\mathrm{Erf}(\cdot)$. Sampling many times u_A and u_B uniformly in the unit interval $(0,1)$, i.e. $u_A, u_B \sim U(0,1)$, and substituting these sampled values in expressions (19) and (20), we obtain simulations of random variables A and B, respectively.

5 Procedure design

In the previous section, we have taken advantage of PME to assign reliable PDFs to model inputs A and B (see expressions (13) and (14), respectively). Taking into account the relations (15) and (16), these PDFs, f_A and f_B, depend on parameters $\left\{\lambda_1^A, \lambda_2^A, a_2\right\}$ and $\left\{\lambda_1^B, \lambda_2^B, b_2\right\}$, respectively. In this section, we design a computational procedure to determine these parameters so that the PDF, f_{X_n}, which according to (6) depends on A and B, matches, as much as possible, the PDFs constructed via PME in Section 3 of sampled data at the time instants $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$.

To seek the parameters λ_1^A, λ_2^A, a_2, λ_1^B, λ_2^B and b_2, an optimization algorithm will be applied. This technique consists of comparing, over various iterations, sets of admissible parameters $(\lambda_1^A, \lambda_2^A, a_2, \lambda_1^B, \lambda_2^B, b_2)$ until an optimum or a satisfactory set is found [31].

To compare sets of admissible parameters, a suitable criterion, which is enclosed in a *fitness function*, is required. In our case, given a set of parameters $(\lambda_1^A, \lambda_2^A, a_2, \lambda_1^B, \lambda_2^B, b_2)$, we have chosen the sum of certain local errors, $E_{\tilde{n}}$, $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$, which are defined in terms of the absolute differences between the PDF, $f_{X_{\tilde{n}}}$, given in (6) and the PDF, $\tilde{f}_{\tilde{n}}$, assigned to sampled data given in equation (10) and Table 2.

Down below, we shall describe through several steps the construction of the fitness function, $FF(s)$, for a given set of parameters $s = (\lambda_1^A, \lambda_2^A, a_2, \lambda_1^B, \lambda_2^B, b_2)$.

Step 1: Compute the values of λ_0^A and λ_0^B defined by equations (15) and (16), respectively.

Step 2: Obtain $M = 10000$ samples of $u_A, u_B \sim U(0, 1)$ and substitute them in equations (19) and (20) to sampling values a and b of random variables A and B, respectively.

Step 3: Define the mesh of $N + 1$ nodes over the interval $[0, H]$,

$$\hat{x} := \{x_i\} := \left\{ \frac{iH}{N} \right\}_{i=0}^{N},$$

being $H < +\infty$ an upper bound of the random variable defined by equation (10) at $\tilde{n} = 48$. In our application we will take $N = 500$ and $H = 8000$ (see panel corresponding $\tilde{n} = 48$ in Figure 1).

Step 4: Fix $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$ and x_i defined in Step 3. Substitute the M simulations (a, b) of the random vector (A, B) generated in Step 2 in the expectation argument of (6), i.e., in the expression

$$f_{X_0}\left(\frac{x_i(a-1)}{a^{\tilde{n}}(a-1) - bx_i(a^{\tilde{n}}-1)}\right)\left|\frac{(a-1)^2 a^{\tilde{n}}}{\left(a^{\tilde{n}}(a-1) - bx_i(a^{\tilde{n}}-1)\right)^2}\right|, \tag{21}$$

Thus, for each $\tilde{n}$, M curves, along the mesh $\hat{x}$, are generated.

Step 5: For each day $\tilde{n}$, compute the average of the M curves generated in Step 4. Then, according to (6) an approximation of the PDF $f_{X_{\tilde{n}}}$ evaluated in $\hat{x}$ is obtained.

Step 6: For each day $\tilde{n}$, evaluate in the mesh $\hat{x}$ the PDF, $\tilde{f}_{\tilde{n}}$, of sampled data defined by equation (10) and Table 2.

Step 7: For each day $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$, compute the error

$$E_{\tilde{n}} = \frac{\sum_{i=0}^{N} \left| f_{X_{\tilde{n}}}(x_i) - \tilde{f}_{\tilde{n}}(x_i) \right|}{\sum_{i=0}^{N} \tilde{f}_{\tilde{n}}(x_i)}.$$

Step 8: The output of the fitness function, named fitness, is given by

$$E = E_0 + E_{16} + E_{30} + E_{33} + E_{43} + E_{48}.$$

It is important to remark that $E_0 = 0$, since by construction we have taken $f_{X_0} = \tilde{f}_0$, see Section 3.

Using an optimization algorithm, we can find out the vector $s = (\lambda_1^A, \lambda_2^A, a_2, \lambda_1^B, \lambda_2^B, b_2)$ that minimizes the fitness E, i.e. a set of parameters such that $f_{X_{\tilde{n}}}$ and $\tilde{f}_{\tilde{n}}$, are close at the time instants $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$.

The optimization algorithm used in this contribution to minimize FF is a bioinspired algorithm named *Particle Swarm Optimization* (PSO). These kind of algorithms are inspired by biological behavior of certain species. In this case, PSO represents the movement of a swarm of birds exploring new areas to find food. In each iteration all the birds of the swarm, change their position according to balance of its particular best position and the global best position of the swarm, [32].

6 Results

This section is aimed at seeking the values $\lambda_1^A, \lambda_2^A, a_2, \lambda_1^B, \lambda_2^B$ and b_2 that minimize the fitness function, FF, described in the foregoing section through Steps $1 - 8$. Minimizing FF, we guarantee that the PDF of the randomized discrete logistic model (6), at the time instants $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$, approximates with the PDF of sampled data described in (10) and Table 2.

As it has been explained in Section 5, PSO algorithm is applied to find out the best set of parameters $s = (\lambda_1^A, \lambda_2^A, a_2, \lambda_1^B, \lambda_2^B b_2)$ that minimizes FF. We consider a swarm made up of 200 particles (birds), and during 90 iterations, the particles change their positions. In other words, our optimization algorithm requires 90 iterations.

Using the MATLAB function *particleswarm* with 200 elements and 90 iterations, we proceed to find out the best set of parameters that minimizes FF. This procedure requires about 3 hours to reach a suitable solution with an Intel Core i7 7700HQ and 16Gb of RAM. The best set of parameters and their respective fitness are collected in Table 3. Notice that the values of λ_2^A and λ_2^B are positive as required in (15) and (16).

Notice that the upper bound, b_2, of the domain of the random variable B is close to zero, $b_2 = 9.365588 \cdot 10^{-6}$. This numerical result is in full agreement with expression (3) that relates the parameter B, appearing in the discrete logistic model (1), and the parameters a and b involved in the formulation of continuous logistic model (2). On the one hand, the numerator of B in expression (3), $e^a - 1$, is small since the random variable $A = e^a$ takes values similar to 1. On the other hand, the denominator of B in (3) is given by the parameter b of the logistic model (2), defining the carrying capacity, i.e. the maximum volume the tumor can reach. From Table 1, we can see that the maximum sampled volume is 2558.6 mm^3, and according to the trend of sampled data, it is expected the carrying capacity, b, will be greater than 2558.6 mm^3. As a consequence, random variable B takes values close to zero.

Parameters	Values
λ_1^A	-2038.1233
λ_2^A	919.6327
a_2	1.11057
λ_1^B	90.5919
λ_2^B	196.5526
b_2	9.365588e-06
Fitness	1.582584

Table 3: Values of parameters $\lambda_1^A, \lambda_2^A, a_2, \lambda_1^B, \lambda_2^B, b_2$ that minimize the fitness function FF using Particle Swarm Optimization algorithm.

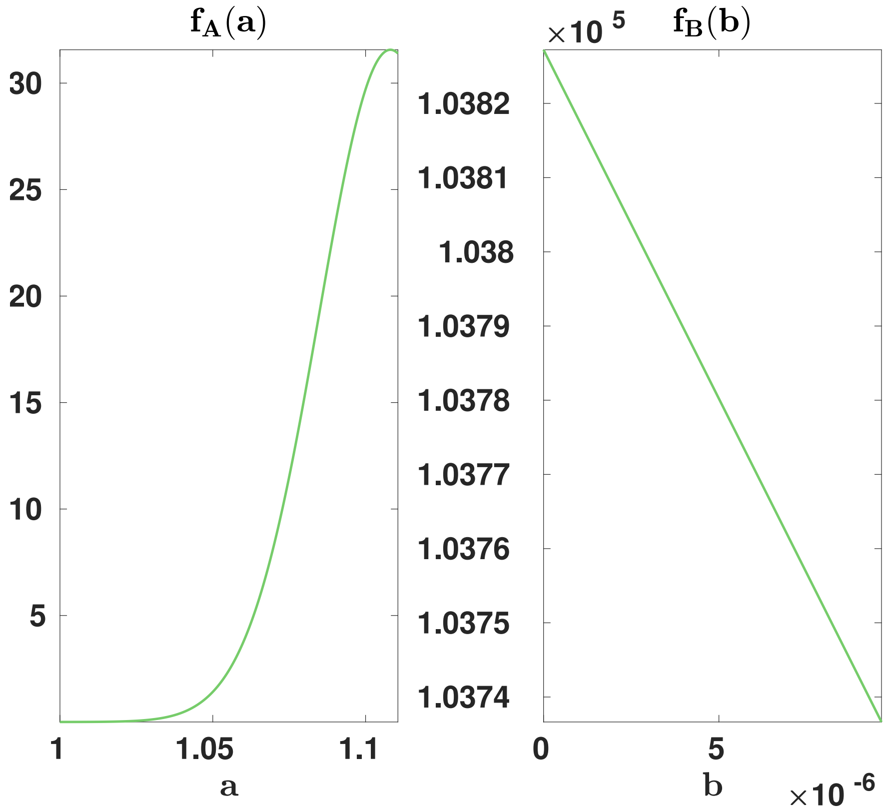

Figure 2: Probability density functions of random model parameters A and B of the randomized discrete logistic model described by equations (13) and (14), respectively.

A graphical representation of the PDFs of random variables A and B described by expressions (13) and (14), respectively, are plotted in Fig. 2.

To better compare the obtained results, in Fig. 3 we show, at the days $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$, the PDF of the randomized discrete logistic model described in equation (6) (blue lines) and the PDF of sampled data described in expression (10) and Table 2 (red dashed lines). We can observe that there is a good agreement between both PDFs at every value of $\tilde{n}$. This confirms the goodness of the fitting procedure.

In Fig. 4, we have plotted the PDF of the randomized discrete logistic model given by equation (6) for $n \in \{0, 1, \dots 50\}$. The red points represent the sampled data (given in column $\tilde{m}_{1,\tilde{n}}$ of Table 1) and green points are the means or expectations obtained via the PDFs (blue curves).

7 Conclusions

In this contribution a probabilistic logistic-type model to describe the growth of breast tumor volume has been presented. A key aspect to treat model uncertainties has been the allocation of reliable distributions to model parameters. To handle this important issue, we have devised a computational method which takes advantage of the principle of maximum entropy. A relevant aspect of our approach is that we fit the model to real data taking into account the probabilistic information via the probability density functions assigned and computed to sampled data and output model. This is a distinctive feature of our study with respect to alternative methods that perform the fitting by means of punctual statistics like the expectation. In this manner, uncertainty quantification analysis in the stochastic model is more informative. The uncertainty

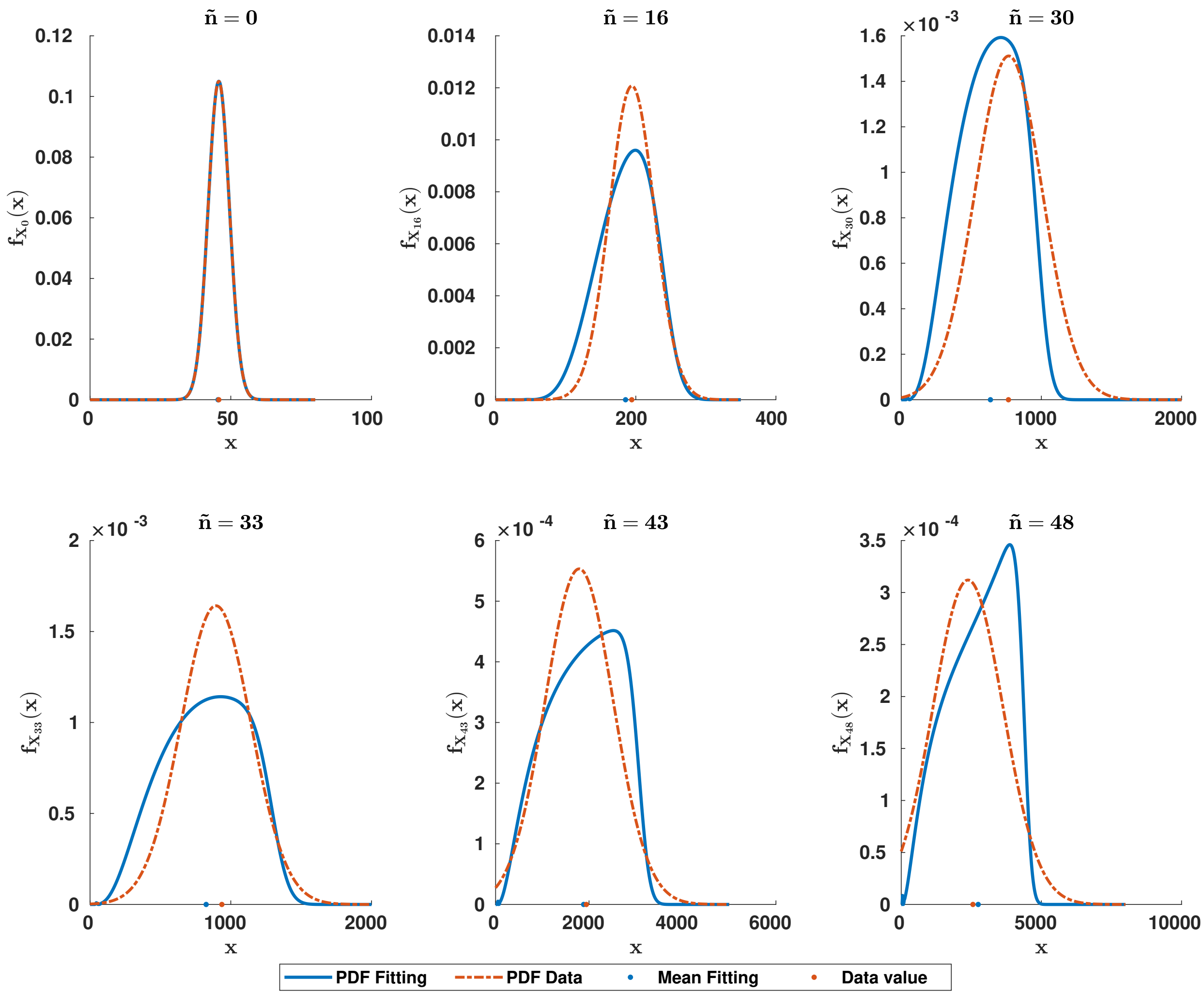

Figure 3: Comparison between the PDF of each sampled data (red dashed lines) and the PDF of the fitting randomized logistic model (blue lines) at the days $\tilde{n} \in \{0, 16, 30, 33, 43, 48\}$. In the horizontal axis of each panel, the red point represents the sampled data (it corresponds to column $\tilde{m}_{1,\tilde{n}}$ in Table 1) and the blue point represents the mean or expectation obtained via the PDF (blue curves).

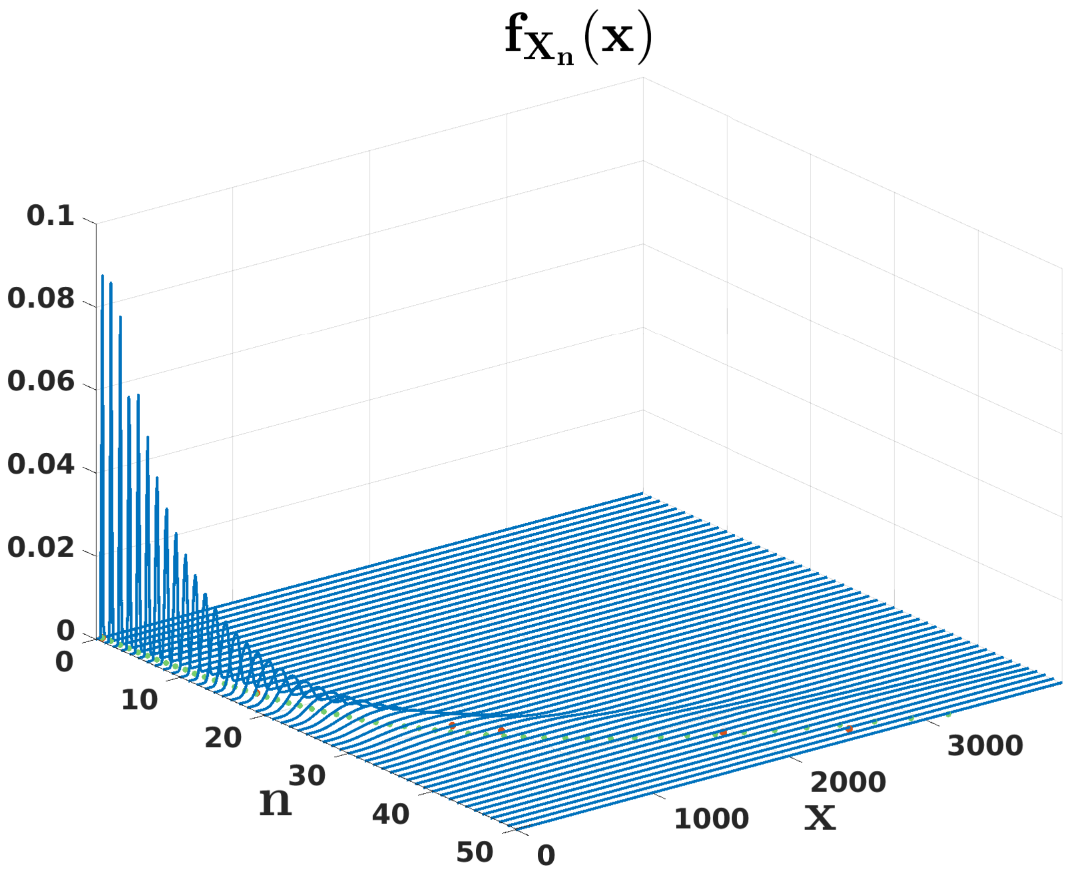

Figure 4: Representation of the PDF, f_{X_n}, of the random discrete logistic model (6) for different days $n = 0, 1, \ldots, 50$. Red points represent the sampled values of tumor volume described in the column $\tilde{m}_{1,\tilde{n}}$ of Table 1 and green points represent the mean of the distribution defined by the blue curves.

quantification technique proposed in this paper may be applied to other models where randomness in data and model parameters play a key role [33, 34].

Conflict of Interest

The authors declare that they do not have any conflict of interest.

Acknowledgements

This work has been supported by the Spanish Ministerio de Economía, Industria y Competitividad (MINECO), the Agencia Estatal de Investigación (AEI), and Fondo Europeo de Desarrollo Regional (FEDER UE) Grants MTM2017-89664-P and RTI2018-095180-B-I00.

References

[1] D. Delen, G. Walker, and A. Kadam", "Predicting breast cancer survivability: a comparison of three data mining methods," Artificial Intelligence in Medicine, vol. 34, no. 2, pp. 113 – 127, 2005. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0933365704001010

[2] H.-C. Wei, "Mathematical modeling of tumor growth: the mcf-7 breast cancer cell line," vol. 16, no. mbe-16-06-325, p. 6512, 2019. [Online]. Available: https://doi.org/10.3934/mbe.2019325.

[3] N. Bellomo and E. de Angelis, Selected Topics in Cancer Modeling: Genesis, Evolution, Immune Competition, and Therapy. Springer Science & Business Media, 2008.

[4] N. Boyd, L. Martin, A. Gunasekara, O. Melnichouk, G. Maudsley, C. Peressotti, M. Yaffe, and S. Minkin, "Mammographic density and breast cancer risk: evaluation of a novel method of measuring breast tissue volumes," Cancer Epidemiology and Prevention Biomarkers, vol. 18, no. 6, pp. 1754–1762, 2009. [Online]. Available: https://doi.org/10.1158/1055-9965.EPI-09-0107

[5] J. E. Joy, E. E. Penhoet, and D. B. Petitti, Saving women's Lives: Strategies for Improving Breast Cancer Detection and Diagnosis. National Academies Press (US), 2005.

[6] J. V. Michalowicz, J. M. Nichols, and F. Bucholtz, Handbook of Differential Entropy. CRC Press, 2013.

[7] C. Burgos-Simón, N. García-Medina, D. Martínez-Rodríguez, and R.-J. Villanueva, "Mathematical modeling of the dynamics of the bladder cancer and the immune response applied to a patient: Evolution and short-term prediction," Mathematical Methods in the Applied Sciences, vol. 42, no. 17, pp. 5746–5757, 2019. [Online]. Available: https://doi.org/10.1002/mma.5536

[8] J. Manimaran, L. Shangerganesh, A. Debbouche, and V. Antonov, "Numerical solutions for time-fractional cancer invasion system with nonlocal diffusion," Frontiers in Physics, vol. 7, p. 93, 2019. [Online]. Available: https://doi.org/10.3389/fphy.2019.00093

[9] S. Chakraborty, A. Debbouche, and P. K. Roy, "A mathematical modelling for treatment of hpv associated cervical cancer: Nk and effector t cell based control study," Nonlinear Studies, vol. 27, no. 2, pp. 325–336, 2020. [Online]. Available: http://www.nonlinearstudies.com/index.php/nonlinear/article/view/1842

[10] D.-A. Botesteanu, S. Lipkowitz, J.-M. Lee, and D. Levy, “Mathematical models of breast and ovarian cancers,” Wiley Interdisciplinary Reviews: Systems Biology and Medicine, vol. 8, no. 4, pp. 337–362, 2016. [Online]. Available: https://doi.org/10.1002/wsbm.1343.

[11] N. C. Atuegwu, L. R. Arlinghaus, X. Li, A. B. Chakravarthy, V. G. Abramson, M. E. Sanders, and T. E. Yankeelov, “Parameterizing the logistic model of tumor growth by DW-MRI and DCE-MRI data to predict treatment response and changes in breast cancer cellularity during neoadjuvant chemotherapy,” Translational oncology, vol. 6, no. 3, pp. 256–264, 2013. [Online]. Available: https://doi.org/10.1593/tlo.13130

[12] H. Enderling, M. A. Chaplain, A. R. Anderson, and J. S. Vaidya, “A mathematical model of breast cancer development, local treatment and recurrence,” Journal of theoretical biology, vol. 246, no. 2, pp. 245–259, 2007. [Online]. Available: https://doi.org/10.1016/j.jtbi.2006.12.010

[13] E. Allen, Modeling with Itô Stochastic Differential Equations. Springer Science & Business Media, 2007, vol. 22.

[14] D. Nualart, The Malliavin Calculus and Related Topics. Springer, 2006, vol. 1995.

[15] R. C. Smith, Uncertainly Quantification: Theory, Implementation, and Applications, ser. Computational Science & Engineering. SIAM, 2013.

[16] M.-C. Casabán, J.-C. Cortés, J.-V. Romero, and M.-D. Roselló, “Probabilistic solution of random SI-type epidemiological models using the Random Variable Transformation technique,” Communications in Nonlinear Science and Numerical Simulation, vol. 24, pp. 86–97, 2015. [Online]. Available: https://doi.org/10.1016/j.cnsns.2014.12.016

[17] M.-C. Casabán, J.-C. Cortés, A. Navarro-Quiles, J.-V. Romero, M.-D. Roselló, and R.-J. Villanueva, “A comprehensive probabilistic solution of random SIS-type epidemiological models using the random variable transformation technique,” Communications in Nonlinear Science and Numerical Simulation, vol. 32, pp. 199–210, 2016. [Online]. Available: https://doi.org/10.1016/j.cnsns.2015.08.009

[18] M. Ezechiáš and T. Cajthaml, “Receptor partial agonism and method to express receptor partial activation with respect to novel full logistic model of mixture toxicology,” Toxicology, vol. 393, pp. 26–33, 2018.

[19] F. Brauer, C. Castillo-Chavez, and C. Castillo-Chavez, Mathematical Models in Population Biology and Epidemiology. Springer, 2012, vol. 2.

[20] C. Salas-Eljatib, A. Fuentes-Ramirez, T. G. Gregoire, A. Altamirano, and V. Yaitul, “A study on the effects of unbalanced data when fitting logistic regression models in ecology,” Ecological Indicators, vol. 85, pp. 502–508, 2018.

[21] V. G. Vaidya and F. J. Alexandro Jr, “Evaluation of some mathematical models for tumor growth,” International journal of bio-medical computing, vol. 13, no. 1, pp. 19–35, 1982. [Online]. Available: https://doi.org/10.1016/0020-7101(82)90048-4

[22] E. C. Pielou, “An introduction to mathematical ecology.” An introduction to mathematical ecology., 1969.

[23] J. G. Skellam, “Random dispersal in theoretical populations,” Biometrika, vol. 38, no. 1/2, pp. 196–218, 1951.

[24] P.-F. Verhulst, “Notice sur la loi que la population suit dans son accroissement,” Corresp. Math. Phys., vol. 10, pp. 113–126, 1838.

[25] T. T. Soong, “Random Differential Equations in Science and Engineering,” 1973.

[26] J.-C. Cortés, A. Navarro-Quiles, J.-V. Romero, and M.-D. Roselló, “(cmmse2018 paper) solving the random Pielou logistic equation with the random variable transformation technique: Theory and applications,” Mathematical Methods in the Applied Sciences, vol. 42, no. 17, pp. 5708–5717, 2019. [Online]. Available: https://doi.org/10.1002/mma.5440

[27] A. Worschech, N. Chen, A. Y. Yong, Q. Zhang, Z. Pos, S. Weibel, V. Raab, M. Sabatino, A. Monaco, H. Liu et al., “Systemic treatment of xenografts with vaccinia virus glv-1h68 reveals the immunologic facet of oncolytic therapy,” BMC genomics, vol. 10, no. 1, p. 301, 2009. [Online]. Available: https://doi.org/10.1186/1471-2164-10-301

[28] H. Caysa, S. Hoffmann, J. Luetzkendorf, L. P. Mueller, S. Unverzagt, K. Mäder, and T. Mueller, “Monitoring of xenograft tumor growth and response to chemotherapy by non-invasive in vivo multispectral fluorescence imaging,” PLoS One, vol. 7, no. 10, p. e47927, 2012.

[29] The MathWorks Inc. (2020) Particle swarm optimization. [Online]. Available: https://es.mathworks.com/help/gads/particleswarm.html

[30] L. Devroye, “Nonuniform random variate generation,” Handbooks in operations research and management science, vol. 13, pp. 83–121, 2006.

[31] T. Weise, “Global optimization algorithms-theory and application,” Self-Published Thomas Weise, 2009.

[32] N. Khemka and C. Jacob, “Exploratory toolkit for evolutionary and swarm-based optimization,” The Mathematica Journal, vol. 11, no. 3, pp. 376–391, feb 2010. [Online]. Available: https://doi.org/10.3888/tmj.11.3-5

[33] Y. Khan, H. Vazquez-Leal, and Q. Wu, “An efficient iterated method for mathematical biology model,” Neural Computing and Applications, vol. 23, no. 3-4, pp. 677–682, 2013.

[34] E. F. D. Goufo, Y. Khan, and Q. A. Chaudhry, “Hiv and shifting epicenters for covid-19, an alert for some countries,” Chaos, Solitons & Fractals, vol. 139, p. 110030, 2020.