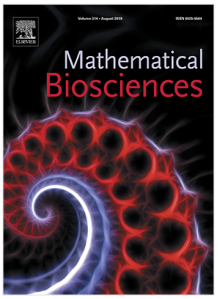
Journal Pre-proof



Inference on an heteroscedastic Gompertz tumor growth model

G. Albano, V. Giorno, P. Román-Román, S. Román-Román,
J.J. Serrano-Pérez, F. Torres-Ruiz

PII: S0025-5564(20)30100-0
DOI: https://doi.org/10.1016/j.mbs.2020.108428
Reference: MBS 108428

To appear in: *Mathematical Biosciences*

Received date : 18 April 2020
Revised date : 16 July 2020
Accepted date : 19 July 2020

Please cite this article as: G. Albano, V. Giorno, P. Román-Román et al., Inference on an heteroscedastic Gompertz tumor growth model, *Mathematical Biosciences* (2020), doi: https://doi.org/10.1016/j.mbs.2020.108428.

Paper: **Inference on an heteroscedastic Gompertz tumor growth model**

Authors: **Albano, Giorno, Roman, Roman, Serrano and Torres**

Highlights

- Modeling tumor dynamics under the effect of anti-angiogenic therapies
- Modeling anti-angigenic and pro-apoptotic therapies
- Inference on the effectiveness of the treatement
- Monitoring by means of test hypothesis the effectiveness of an experimental therapy.

Inference on an heteroscedastic Gompertz tumor growth model

G. Albano[a,*], V. Giorno[b], P. Román-Román[c,d], S. Román-Román[e], J.J. Serrano-Pérez[c], F. Torres-Ruiz[c,d]

[a]*Dipartimento di Studi Politici e Sociali, Università di Salerno, Italy*
[b]*Dipartimento di Informatica, Università di Salerno, Italy*
[c]*Departamento de Estadística e Investigación Operativa, Universidad de Granada, Spain*
[d]*Instituto de Matemáticas (IEMath-GR), Universidad de Granada, Spain*
[e]*Dép. de Recherche Translationnelle, Institut Curie, France*

Abstract

We consider a non homogeneous Gompertz diffusion process whose parameters are modified by generally time-dependent exogenous factors included in the infinitesimal moments. The proposed model is able to describe tumor dynamics under the effect of anti-proliferative and/or cell death-induced therapies. We assume that such therapies can modify also the infinitesimal variance of the diffusion process. An estimation procedure, based on a control group and two treated groups, is proposed to infer the model by estimating the constant parameters and the time-dependent terms. Moreover, several concatenated hypothesis tests are considered in order to confirm or reject the need to include time-dependent functions in the infinitesimal moments. Simulations are provided to evaluate the efficiency of the suggested procedures and to validate the testing hypothesis. Finally, an application to real data is considered.

Key words: Tumor growth, anti-proliferative and cell death-induced therapies, modified Gompertz diffusion process, inference in diffusion processes, bootstrap tests.

*Corresponding author
Email addresses: pialbano@unisa.it (G. Albano), giorno@unisa.it (V. Giorno), proman@ugr.es (P. Román-Román), sergio.Roman-Roman@curie.fr (S. Román-Román), jjserra@ugr.es (J.J. Serrano-Pérez), fdeasis@ugr.es (F. Torres-Ruiz)

1. Introduction

Diffusion processes are widely used in the literature to describe phenomena in a lot of fields, ranging from economics [1, 2] to biology [3, 4]. Concerning the tumor growth modeling, many efforts have been devoted in the last years since nowadays the cancer represents one of the main causes of death in our society. Further, the availability of modern diagnostic and prognostic methodologies allows to build ever more faithful models, giving useful insights into the dynamics of such disease [5–7]. On the other hand, the mathematical tractability of the model must be taken into account because it allows to better handle explicit solutions of the involved dynamics. In this context, Gompertz growth model seems successfully overcome the "trade off" between these two aspects. Indeed, it is widely accepted that such model is able to capture dynamics of solid tumors and several models based on this growth have been proposed by looking at deterministic and stochastic behaviors [8–13].

The Gompertz curve belongs to the Richards family of sigmoidal growth models, along with familiar models such as the negative exponential, the logistic, and the Bertalanffy [14]. These curves, although born in deterministic contexts, have been generalized to include stochastic effects aimed at bridging the gaps that often exist between experimental data and theoretical results. Concerning the stochastic version of the Gompertz growth, in the literature various contributions can be found concerning both theoretical probabilistic properties and statistical characteristics [15–17].

Recently, the attention has been focused on non-homogeneous Gompertz diffusion process describing the effect of some exogenous generally time-dependent factors. In preclinical tumor growth studies it is useful to understand as experimental therapies can modify the natural cancer cells' growth rates. A modified Gompertz equation is considered in Cabrales et al. [18] to describe tumor responses to electrochemical treatments and the possible decay of solutions is investigated both from theoretical and numerical points of view. In [19] a control approach to predict an optimal drug dosage shrinking the cancer tumor-cell

population was proposed. The predictive control problem is hence based on the difference between the probability density function and the desired probability density function calculated at each time instant.

In the mathematical framework, an untreated tumor volume can be modeled by an homogeneous Gompertz stochastic process. In order to include the effect of an anti-angiogenic therapy in [15, 20, 21] the infinitesimal moments of the homogeneous process were modified by introducing suitable continuous time-dependent functions modeling the exogenous factors; in such a way, the therapy presence leads to a time non homogeneous stochastic process. A statistical approach was also proposed in [22–24] to fit the modifications in the natural growth rates due to several therapies.

1.1. Motivations and plan of the paper

In the present paper we provide a natural generalization of the results provide in [23] and [24]. Specifically, we assume that the two applied therapies (one of anti-proliferative and the other inducing the cancer cells death) are generally able to modify both the drift and the infinitesimal variance of the process making it time-dependent. We propose a statistical methodology to estimate the natural tumor rates and to fit the exogenous term including also the infinitesimal variance of the resulting process. The procedure uses a control group $\mathcal{G}$ described by a homogeneous Gompertz diffusion process and two treated groups $\mathcal{G}_1$ and $\mathcal{G}_2$ described by two time non-homogeneous processes. The group $\mathcal{G}_1$ is assumed to be treated with a therapy (anti-proliferative or cell death-induced) while the group $\mathcal{G}_2$ is treated the same therapy of $\mathcal{G}_1$ together with an other therapy of the other type. Further, we consider the case in which the two groups are characterized by two generally different time-dependent infinitesimal variance.

The procedure works as follows. In a first step, the control group $\mathcal{G}$ is used to estimate the constant parameters included in the homogeneous process, whereas in the subsequent steps the treated groups, $\mathcal{G}_1$ and $\mathcal{G}_2$, are used to fit the unknown time-dependent functions describing the effect of the therapies. Precisely, via suitable mathematical relations, the group $\mathcal{G}_1$ is used to fit the

function related to the single therapy applied in it and the effect that this therapy has on the infinitesimal variance. Then, the group $\mathcal{G}_2$ is used to fit the second therapy and its effect on the infinitesimal variance.

Moreover, a bootstrap testing procedure able to evaluate the time dependence of the exogenous factors is provided. Essentially, it is directed to establish if the real effect of the various applied therapies has a known functional form. In particular the proposed test is then used to establish if the therapy effect is null or constant.

We point out that both the estimation procedure and the testing hypothesis on the exogenous factors related to the stochastic diffusion processes are of interest in various applicative and theoretical contexts [25–27].

The plan of the paper is the following. In Section 2 the model is introduced and its probability distribution and some statistical characteristics are derived. In Section 3 the procedure to estimate the parameters and to fit the unknown functions is proposed. Various simulated-based examples are given to validate the fitting procedure. In Section 4 the hypothesis test procedure is provided and several cases of particular interest in biological context are considered. In particular, some concatenated tests are performed to evaluate the constant/null effect of the therapy on the rates and on the infinitesimal variance. Finally, in Section 5 an application to real data is provided to study the combined effect of Carboplatin and Taxol in ovarian cancer.

2. The model

In the mathematical framework, an untreated tumor volume can be modeled by an homogeneous Gompertz stochastic process defined in $\mathbb{R}^+$ with infinitesimal moments

$$\begin{aligned} A_1(x) &= \alpha x - \beta x \log x, \\ A_2(x) &= \sigma^2 x^2, \end{aligned} \tag{1}$$

where α, β and σ are positive constants. The parameters α and β describe the cell's growth and death rates, respectively, σ is related to more or less intense

environmental fluctuations introduced to justify discrepancies between clinical data and theoretical predictions that quite often are detected.

Our approach for including the effect of an anti-angiogenic therapy consists to modify the infinitesimal moments (1) by introducing suitable continuous time-dependent functions modeling the exogenous factors; in such a way, the therapy presence leads to a time non homogeneous stochastic process. Precisely, let $\{X(t) : t \geq t_0\}$ with $t_0 \geq 0$ be a stochastic process in $\mathbb{R}^+$ and satisfying the following stochastic differential equation (SDE)

$$\begin{aligned} dX(t) &= \{(\alpha - C(t)) - (\beta - D(t)) \ln X(t)\} X(t)\, dt + \sigma \sqrt{V(t)} X(t)\, dW(t), \\ X(t_0) &= X_0. \end{aligned} \tag{2}$$

Here, as in (1), α, β and σ are positive constants, while $C(t), D(t)$ and $V(t)$ are functions in $C_1[t_0, +\infty)$ with $V(t) > 0$ for all $t \geq t_0$, X_0 is a random variable describing the initial state of the process, and $W(t)$ is a standard Wiener process independent from X_0 for $t \geq t_0$. In the model setting, $C(t)$ represents tumor regression rate due to the therapy and has the same dimension as parameter α, while the function $D(t)$ modifies the death rate β of the process (1) in $\beta - D(t)$. From a biological point of view, the function $C(t)$ describes the effect of an anti-proliferative therapy, that is, able to modify the natural birth rate of cancer cells, while $D(t)$ describes the effect of cell death-induced therapy (see [15, 23]). Clearly $C(t)$ successfully applied when it assumes positive values, while $D(t)$ is effective when it is a negative function. Further it would be desirable to have small values of the function $V(t)$, describing fluctuations in the tumor volume. Anyway, in experimental studies the effectiveness of a therapy has to be tested, so we assume that the functions $C(t)$, $D(t)$ have real values and $V(t) > 0$.

The aim of this paper is to model the combined effect of two therapies, one anti-proliferative and the other that induces the death of cancer cells. In this sense, model (2) can be viewed as a modification of model (1) after transforming its infinitesimal moments by introducing the functions $C(t)$, $D(t)$ and $V(t)$.

By considering

$$Z(t) = f(X(t), t) = k(t)\, \ln X(t), \tag{3}$$

where

$$k(t) = \exp\left(\int^t [\beta - D(s)]\, ds \right),$$

and by applying Itô's Lemma, we can transform process $X(t)$ into a non-homogeneous Wiener process $Z(t)$ described by the following SDE:

$$dZ(t) = a(t)\, dt + b(t)\, dW(t), \qquad Z(t_0) = Z_0, \tag{4}$$

with

$$a(t) = k(t) \left[\alpha - C(t) - \frac{\sigma^2 V(t)}{2} \right], \qquad b(t) = \sigma \sqrt{V(t)}\, k(t),$$

whose solution is

$$Z(t) = Z_0 + \int_{t_0}^t a(s) ds + \int_{t_0}^t b(s)\, dW(s).$$

Finally, undoing the change (3), we obtain

$$X(t) = \exp\left\{ \frac{1}{k(t)} \left[k(t_0) \ln X_0 + \int_{t_0}^t a(s) ds + \int_{t_0}^t b(s)\, dW(s) \right] \right\}.$$

2.1. Distribution of the process

From (4), and if Z_0 is a degenerate random variable, i.e. $P(Z_0 = z_0) = 1$, with $z_0 \in \mathbb{R}$ or normally distributed, i.e. $Z_0 \sim N_1[\mu_0, \sigma_0^2]$, then $Z(t)$ is a Gaussian process, so, $\forall\, n \in \mathbb{N}$ and $t_1 < \cdots < t_n$, vector $(Z(t_1), \ldots, Z(t_n))^T$ has a n-dimensional normal distribution $N_n[\boldsymbol{\varepsilon}, \boldsymbol{\Sigma}]$, where the components of vector $\boldsymbol{\varepsilon}$ and matrix $\boldsymbol{\Sigma}$ are

$$\varepsilon_i = E[Z_0] + \int_{t_0}^{t_i} k(s) \left[\alpha - C(s) - \frac{\sigma^2 V(s)}{2} \right] ds, \quad i = 1, \ldots, n$$

and

$$\sigma_{ij} = Var[Z_0] + \sigma^2 \int_{t_0}^{\min(t_i, t_j)} k^2(s) V(s)\, ds, \quad i, j = 1, \ldots, n,$$

respectively.

Therefore, by (3) all the finite-dimensional distributions of the process $X(t)$ are lognormal; specifically, $\forall\, n \in \mathbb{N}$,

$$(X(t_1), \ldots, X(t_n))^T \sim \Lambda_n[\boldsymbol{\xi}, \boldsymbol{\Delta}], \tag{5}$$

where $\xi_i = \dfrac{\varepsilon_i}{k(t_i)}$ and $\delta_{ij} = \dfrac{\sigma_{ij}}{k(t_i)k(t_j)}$, $i, j = 1 \ldots, n$, are the components of $\boldsymbol{\xi}$ and $\boldsymbol{\Delta}$, respectively.

In particular, by considering $X_0 \sim \Lambda_1[\mu_0; \sigma_0^2]$, we have

$$X(t) \sim \Lambda_1 \left[M^*(t|\mu_0, t_0); V^*(t|\sigma_0^2, t_0)\right],$$

where, for $\tau < t$,

$$M^*(t|u, \tau) = u\bar{k}(t|\tau) + \int_\tau^t \left(\alpha - C(s) - \frac{\sigma^2 V(s)}{2}\right) \bar{k}(t|s)\, ds,$$

and

$$V^*(t|u, \tau) = u\bar{k}^2(t|\tau) + \int_\tau^t \sigma^2 V(s) \bar{k}^2(t|s)\, ds,$$

with $\bar{k}(t|\tau) = k(\tau)/k(t)$.

In the following we will assume that X_0 is a degenerate random variable in x_0. This assumption is quite common in the context of tumor growth since the variable of interest is usually the relative volume of the tumor, and $x_0 = 1$ is the relative volume at the detection of the tumor. So, we will assume $P[X_0 = x_0] = 1$. In this case

$$X(t) \sim \Lambda_1 \left[m_1(t); u(t)\right],$$

with

$$m_1(t) = M^*(t|\ln x_0, t_0) \tag{6}$$

and

$$u(t) = V^*(t|0, t_0). \tag{7}$$

So, the mean and the variance functions of $X(t)$ are:

$$E[X(t)] = \exp\left(m_1(t) + \frac{1}{2}u(t)\right),$$
$$Var[X(t)] = \exp\left(2m_1(t) + u(t)\right) \times \left[\exp\left(u(t)\right) - 1\right], \tag{8}$$

respectively.

3. Estimation of the model

In this section we propose a procedure to estimate the parameters α, β, σ, and to approximate the functions $C(t)$, $D(t)$ and $V(t)$ in $[t_0, T]$. To this end, in practice it is necessary to have data from three experimental groups of individuals. Concretely:

- an untreated (control) group, say $\mathcal{G}$,
- a first group, $\mathcal{G}_1$, treated with a single therapy that affects only one of the two rates that model the untreated tumor volume,
- a second group, $\mathcal{G}_2$, treated with two therapies. One of them must be the same therapy applied in $\mathcal{G}_1$, whereas the other one affects the rate not modified in $\mathcal{G}_1$.

The control group is associated to the stochastic process $X(t)$ described by the SDE

$$dX(t) = [\alpha - \beta \ln X(t)]X(t)dt + \sigma X(t)dW(t) \qquad X(t_0) = x_0. \tag{9}$$

Moreover, group $\mathcal{G}_1$ is modeled by a stochastic process $X_1(t)$ for which two cases can be considered:

- $\mathcal{G}_1$ is treated with an anti-proliferative therapy, i.e. mainly affecting cell growth. In this case, $X_1(t)$ follows the SDE

$$dX_1(t) = \{[\alpha - C(t)] - \beta \ln X_1(t)\} X_1(t)dt + \sigma\sqrt{V_1(t)}\, X_1(t)dW(t),\ X_1(t_0) = x_0. \tag{10}$$

- $\mathcal{G}_1$ is treated with a therapy that induces, or mainly induces, the death of cancer cells. Now the SDE followed by $X_1(t)$ is

$$dX_1(t) = \{\alpha - [\beta - D(t)] \ln X_1(t)\} X_1(t)dt + \sigma\sqrt{V_1(t)}\, X_1(t)dW(t),\ X_1(t_0) = x_0. \tag{11}$$

Finally, group $\mathcal{G}_2$ is described by a stochastic process $X_2(t)$ solution of

$$dX_2(t) = \{[\alpha - C(t)] - [\beta - D(t)] \ln X_2(t)\} X_2(t)dt + \sigma\sqrt{V_2(t)}\, X_2(t)dW(t),\ X_2(t_0) = x_0. \tag{12}$$

The basic idea is to use data from the control group to estimate the parameters α, β and σ^2, whereas the treated groups are used to fit the functions $C(t)$, $D(t)$, $V_1(t)$ and $V_2(t)$.

3.1. Some basic expressions

In this subsection we introduce some expressions that are the basis of the estimation procedure developed in the next one.

From (8), we define

$$m_2(t) = \ln E[X(t)] = m_1(t) + \frac{1}{2}u(t),$$

and by considering (6) and (7), after some algebra, the following relationships are obtained:

$$\begin{aligned} C(t) &= \alpha - (\beta - D(t))(m_1(t) + u(t)) - m_1'(t) - \frac{1}{2}u'(t) \\ &= \alpha - (\beta - D(t))(2m_2(t) - m_1(t)) - m_2'(t) \end{aligned} \tag{13}$$

$$D(t) = \beta + \frac{m_1'(t) + \frac{1}{2}u'(t) - \alpha + C(t)}{m_1(t) + u(t)} = \beta + \frac{m_2'(t) - \alpha + C(t)}{2m_2(t) - m_1(t)} \tag{14}$$

$$\begin{aligned} V(t) &= \frac{1}{\sigma^2}(u'(t) + 2(\beta - D(t))u(t)) \\ &= \frac{2}{\sigma^2}[(m_2'(t) - m_1'(t)) + 2(\beta - D(t))(m_2(t) - m_1(t))]. \end{aligned} \tag{15}$$

We point out that the two expressions obtained for $C(t), D(t)$ and $V(t)$ in (13), (14) and (15) respectively, can be alternatively used to fit the functions depending on the behavior of the sampling versions of these functions in real applications.

3.2. The estimation procedure

Let us consider d sample-paths from the control group, observed at the same time instants t_j, $j = 0, \ldots, n-1$, in the interval $[t_0, T]$. Let $\{x_{ij},\, i = 1, \ldots, d;\, j = 0, \ldots, n-1\}$ be the observed values of the sample paths. Moreover, let $\{x_{ij}^{(k)},\, i = 1, \ldots, d_k;\, j = 0, \ldots, n-1\}$ be the values of d_k sample paths from the treated group $\mathcal{G}_k$, $k = 1, 2$, observed at the same previous time instants.

Making use of Equations (13)-(15), and denoting $m_1^{(k)}(t) = E[\ln X_k(t)]$, $m_2^{(k)}(t) = \ln E[X_k(t)]$ and $u^{(k)}(t) = Var[\ln X_k(t)]$, $k = 1, 2$, we can estimate the three models from data provided by the control and the two treated groups. To this end we provide the following stepwise procedure:

- Obtain the maximum likelihood (ML) estimates of α, β and σ^2 by solving the likelihood equation system (18)-(20) in Appendix for $C(t) = D(t) = 0$ and $V(t) = 1$, from the data of group $\mathcal{G}$. Denote by $\widehat{\alpha}$, $\widehat{\beta}$ and $\widehat{\sigma}^2$ such estimates.

- Calculate at each time instant t_j, $j = 0, \ldots, n-1$, the values

$$\widehat{m}_1^{(k)}(t_j) = \bar{y}_j^k,\ \widehat{m}_2^{(k)}(t_j) = \ln(\bar{x}_j^{(k)}),\ \widehat{u}^{(k)}(t_j) = s_j^{2(k)},\ k = 1, 2$$

  where

  - $\bar{y}_j^k$ is the sample mean of the logarithms of the values of the sample paths of the group $\mathcal{G}_k$ $(k = 1, 2)$ at t_j,
  - $\bar{x}_j^{(k)}$ is the sample mean of the values of the sample paths of $\mathcal{G}_k$ $(k = 1, 2)$ at t_j,
  - $s_j^{2(k)}$ is the unbiased sample variance of the logarithms of the values of the sample paths of $\mathcal{G}_k$ $(k = 1, 2)$ at t_j.

- For $k = 1, 2$, approximate the derivatives of $m_1^{(k)}(t)$, $m_2^{(k)}(t)$ and $u^{(k)}(t)$, at t_j from the values obtained in the previous step. Denote by $\widehat{m}_1^{(k)'}(t_j)$, $\widehat{m}_2^{(k)'}(t_j)$ and $\widehat{u}^{(k)'}(t_j)$ the obtained values.

- Estimating $C(t)$, $D(t)$, $V_1(t)$ and $V_2(t)$ as follows:

  – If $\mathcal{G}_1$ is modeled by (10), i.e. it is treated with an anti-proliferative therapy, obtain an initial estimate of $C(t_j)$ and $V_1(t_j)$ by applying the observed data of this group to expressions (13) and (15), with $D(t) = 0$. This leads to

$$\begin{aligned}\widehat{C}_j &= \widehat{\alpha} - \widehat{\beta}\left(\widehat{m}_1^{(1)}(t_j) + \widehat{u}^{(1)}(t_j)\right) - \widehat{m}_1^{(1)'}(t_j) - \frac{1}{2}\widehat{u}^{(1)'}(t_j) \\ &= \widehat{\alpha} - \widehat{\beta}\left(2\widehat{m}_2^{(1)}(t_j) - \widehat{m}_1^{(1)}(t_j)\right) - \widehat{m}_2^{(1)'}(t_j)\end{aligned}$$

  and

$$\begin{aligned}\widehat{V}_{1,j} &= \frac{1}{\widehat{\sigma}^2}\left(\widehat{u}^{(1)'}(t_j) + 2\widehat{\beta}\widehat{u}^{(1)}(t_j)\right) \\ &= \frac{2}{\widehat{\sigma}^2}\left(\widehat{m}_2^{(1)'}(t_j) - \widehat{m}_1^{(1)'}(t_j) + 2\widehat{\beta}\left(\widehat{m}_2^{(1)}(t_j) - \widehat{m}_1^{(1)}(t_j)\right)\right).\end{aligned}$$

  Next, for each t_j, calculate initial estimates of $D(t_j)$ and $V_2(t_j)$ for process $X_2(t)$, by considering (14) and (15) for the data of group $\mathcal{G}_2$ and the previous $\widehat{C}_j$ values. In this way the following values are obtained:

$$\begin{aligned}\widehat{D}_j &= \widehat{\beta} + \frac{\widehat{m}_1^{(2)'}(t_j) + \frac{1}{2}\widehat{u}^{(2)'}(t_j) - \widehat{\alpha} + \widehat{C}_j}{\widehat{m}_1^{(2)}(t_j) + \widehat{u}^{(2)}(t_j)} \\ &= \widehat{\beta} + \frac{\widehat{m}_2^{(2)'}(t_j) - \widehat{\alpha} + \widehat{C}_j}{2\widehat{m}_2^{(2)}(t_j) - \widehat{m}_1^{(2)}(t_j)}\end{aligned}$$

  and

$$\begin{aligned}\widehat{V}_{2,j} &= \frac{1}{\widehat{\sigma}^2}\left(\widehat{u}^{(2)'}(t_j) + 2(\widehat{\beta} - \widehat{D}_j)\widehat{u}^{(2)}(t_j)\right) \\ &= \frac{2}{\widehat{\sigma}^2}\left(\widehat{m}_2^{(2)'}(t_j) - \widehat{m}_1^{(2)'}(t_j) + 2(\widehat{\beta} - \widehat{D}_j)\left(\widehat{m}_2^{(2)}(t_j) - \widehat{m}_1^{(2)}(t_j)\right)\right).\end{aligned}$$

– If $\mathcal{G}_1$ is treated with a therapy that induces the death of cancer cells, i.e. the model (11) is now considered, determine values $\widehat{D}_j$ and $\widehat{V}_{1,j}$ (initial estimates of $D(t_j)$ and $V_1(t_j)$, $j = 0, \ldots, n-1$) from (14) and (15) by considering $C(t) = 0$ and the data of $\mathcal{G}_1$, thus obtaining

$$\begin{aligned}\widehat{D}_j &= \widehat{\beta} + \frac{\widehat{m}_1^{(1)'}(t_j) + \frac{1}{2}\widehat{u}^{(1)'}(t_j) - \widehat{\alpha}}{\widehat{m}_1^{(1)}(t_j) + \widehat{u}^{(1)}(t_j)} \\ &= \widehat{\beta} + \frac{\widehat{m}_2^{(1)'}(t_j) - \widehat{\alpha}}{2\widehat{m}_2^{(1)}(t_j) - \widehat{m}_1^{(1)}(t_j)}\end{aligned}$$

and

$$\begin{aligned}\widehat{V}_{1,j} &= \frac{1}{\widehat{\sigma}^2}\left(\widehat{u}^{(1)'}(t_j) + 2(\widehat{\beta} - \widehat{D}_j)\widehat{u}^{(1)}(t_j)\right) \\ &= \frac{2}{\widehat{\sigma}^2}\left(\widehat{m}_2^{(1)'}(t_j) - \widehat{m}_1^{(1)'}(t_j) + 2(\widehat{\beta} - \widehat{D}_j)\left(\widehat{m}_2^{(1)}(t_j) - \widehat{m}_1^{(1)}(t_j)\right)\right).\end{aligned}$$

Then, for process $X_2(t)$, compute initial estimates of $C(t_j)$ and $V_2(t_j)$, t_j, $j = 0, \ldots, n-1$, from (13) and (15) by taking the data of group $\mathcal{G}_2$ and the values $\widehat{D}_j$ previously estimated. This leads to

$$\begin{aligned}\widehat{C}_j &= \widehat{\alpha} - (\widehat{\beta} - \widehat{D}_j)\left(\widehat{m}_1^{(2)}(t_j) + \widehat{u}^{(2)}(t_j)\right) - \widehat{m}_1^{(2)'}(t_j) - \frac{1}{2}\widehat{u}^{(2)'}(t_j) \\ &= \widehat{\alpha} - (\widehat{\beta} - \widehat{D}_j)\left(2\widehat{m}_2^{(2)}(t_j) - \widehat{m}_1^{(2)}(t_j)\right) - \widehat{m}_2^{(2)'}(t_j)\end{aligned}$$

and

$$\begin{aligned}\widehat{V}_{2,j} &= \frac{1}{\widehat{\sigma}^2}\left(\widehat{u}^{(2)'}(t_j) + 2(\widehat{\beta} - \widehat{D}_j)\widehat{u}^{(2)}(t_j)\right) \\ &= \frac{2}{\widehat{\sigma}^2}\left(\widehat{m}_2^{(2)'}(t_j) - \widehat{m}_1^{(2)'}(t_j) + 2(\widehat{\beta} - \widehat{D}_j)\left(\widehat{m}_2^{(2)}(t_j) - \widehat{m}_1^{(2)}(t_j)\right)\right).\end{aligned}$$

- Obtain $\widehat{C}(t)$, $\widehat{D}(t)$, $\widehat{V}_1(t)$ and $\widehat{V}_2(t)$ as follows:

  – Calculate the final estimated values $\widehat{C}(t_j)$, $\widehat{D}(t_j)$, $\widehat{V}_1(t_j)$ and $\widehat{V}_2(t_j)$ by using local regression of $\widehat{C}_j$, $\widehat{D}_j$, $\widehat{V}_{1,j}$ and $\widehat{V}_{2,j}$ on t_j, respectively.

  – Interpolate, by means of spline functions[1], the data points $(t_j, \widehat{C}(t_j))$,

[1]Since the functions $C(t)$, $D(t)$ and $V(t)$ are sufficiently smooth (they are C^1-class), we use the natural cubic spline interpolation.

$(t_j, \widehat{D}(t_j))$, $(t_j, \widehat{V}_1(t_j))$ and $(t_j, \widehat{V}_2(t_j))$, respectively.

3.3. Simulation-based applications

In order to validate the proposed estimation procedure, we have developed two applications based on simulated data:

- In the former, we consider an untreated group ($\mathcal{G}$), a first group ($\mathcal{G}_1$) treated with an anti-proliferative therapy (so, it is modeled by (10)), and a second group ($\mathcal{G}_2$) that is treated with the same therapy as the first group together with another one inducing the death of cancer cells. This group is modeled by (12).

- In the second case, in addition to the control group ($\mathcal{G}$), we consider a group $\mathcal{G}_1$ treated with a therapy that induces the death of cancer cells (modeled by (11)), whereas $\mathcal{G}_2$ is treated with the same therapy as the first group together with an anti-proliferative therapy, so this group is modeled by (12).

In the two applications the untreated group is modeled by (9) with the same parameters. Table 1 summarizes the parameters and functions considered. The choice of α, β and σ values has been made so that the simulated paths present values similar to real situations. On the other hand, the therapeutical functions in our simulation experiment are in line with [15, 23]. In Application 1 we consider the case in which the group $\mathcal{G}_1$ is treated with an anti-proliferative linear therapy, while the group $\mathcal{G}_2$ is treated with a cell death-induced therapy having a "bump effect" when it is applied and asymptotically reduces of 12% the natural death rate of the tumor. In Application 2 the two therapies are reversed. The infinitesimal variances $V_1(t)$ and $V_2(t)$ involve two lognormal probability density functions since we expect that the variability of the process is greatly influenced by the therapies when they are applied, then they restore to natural values. This assumption is close to what is observed in real situations like the one presented in Section 5.

Table 1: Parameters and functions considered in each example (being $\Lambda_1(t,\mu,\sigma^2)$ the density function of a lognormal distribution $\Lambda_1(\mu,\sigma^2)$).

Group	**Application 1**	**Application 2**
$\mathcal{G}$	$\alpha = 0.5,\ \ \beta = 0.2,\ \ \sigma = 0.01$	
$\mathcal{G}_1$	$C(t) = 0.005\,t$ $V_1(t) = (0.7{+}10\,\Lambda_1(t,3,0.5))^2$	$D(t) = -0.12\,t^2/(50 + t(t-10))$ $V_1(t) = (0.7{+}10\,\Lambda_1(t,3,0.5))^2$
$\mathcal{G}_2$	$D(t) = -0.12\,t^2/(50 + t(t-10))$ $V_2(t) = (0.7{+}15\,\Lambda_1(t,3,0.5))^2$	$C(t) = 0.005\,t$ $V_2(t) = (0.7{+}10\,\Lambda_1(t,3,0.5))^2$

The estimation procedure has been replicated 100 times in both the examples. In each replication, 25 sample paths have been simulated by considering 51 time instants equally spaced in the interval $[0, 50]$. The sample paths have been simulated using the *snssde1d* function from the R package *Sim.DiffProc* [28]. This function allows to simulate the solution of a stochastic differential equation from its discretization by using differents numerical schema as Euler-Maruyama and Milstein among others (see Iacus [29] for details). Further, a degenerate initial distribution at $x_0 = 1$ has been considered. This choice has been made because in real studies on the evolution of tumors (such as the one shown in section 5) the data provided are relative volumes of the tumors.

The estimates in the control group were $\widehat{\alpha} = 0.496477$, $\widehat{\beta} = 0.198469$ and $\widehat{\sigma} = 0.010043$. In Figure 1 the fit of the functions $C(t), V_1(t), D(t)$ and $V_2(t)$ in the models (10) and (12) in Application 1 are plotted on the top. The mean and variances of the processes $X_1(t)$ and $X_2(t)$ along with their fitted versions are also shown on the bottom. The absolute difference functions between the simulated and fitted function are also represented in green. Results related to Application 2 are shown in Figure 2. In both the applications, the procedure provides estimated functions (red lines) very close to the theoretical ones (black lines).

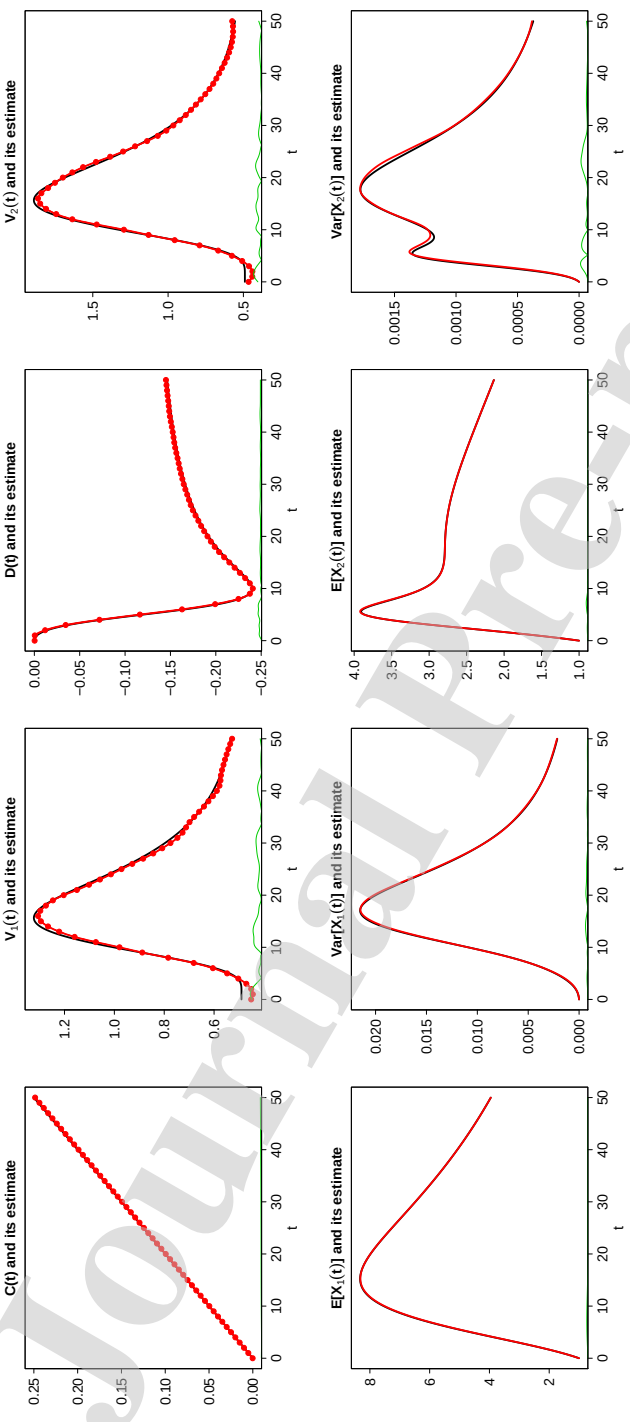


Figure 1: Fit of the functions $C(t), V_1(t), D(t)$ and $V_2(t)$ in the models (10) and (12) in Application 1 are plotted on the top. The mean and variances of the processes $X_1(t)$ and $X_2(t)$ along with their fitted versions (in red) are also shown on the bottom. The absolute difference functions between the simulated and fitted function are represented in green.

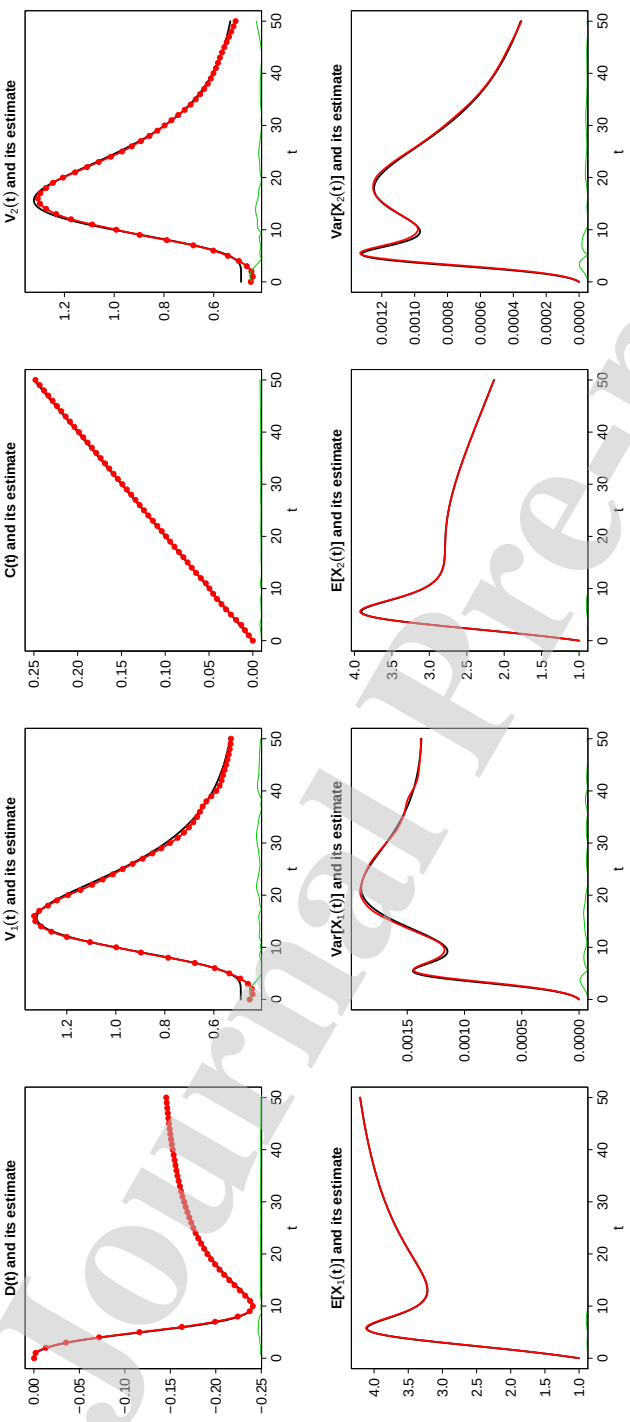


Figure 2: Fit of the functions $D(t), V_1(t), C(t)$ and $V_2(t)$ in the models (11) and (12) in Application 2 are plotted on the top. The mean and variances of the processes $X_1(t)$ and $X_2(t)$ along with their fitted versions (in red) are also shown on the bottom. The absolute difference functions between the simulated and fitted function are represented in green.

Moreover, in order to have a measure of goodness of fit of the obtained estimates, we have considered the following mean squared error

$$MSE(H) = \frac{1}{n}\sum_{j=0}^{n-1}\left(\widehat{H}(t_j) - H(t_j)\right)^2,$$

where H represents one of the C, D, V_1, V_2 functions. Table 2 includes the values obtained for these errors, confirming the closeness between theoretical and estimated functions.

Table 2: Mean squared errors of estimated functions in groups $\mathcal{G}_1$ and $\mathcal{G}_2$ for Applications 1 and 2

Application 1			
Group $\mathcal{G}_1$		Group $\mathcal{G}_2$	
Function	MSE	Function	MSE
$C(t)$	$5.779772e-07$	$D(t)$	$2.637550e-06$
$V_1(t)$	$3.006884e-04$	$V_2(t)$	$4.192476e-04$
$E(X_1(t))$	$1.684190e-04$	$E(X_2(t))$	$2.448863e-05$
$Var(X_1(t))$	$7.953060e-09$	$Var(X_2(t))$	$5.742807e-10$
Application 2			
Group $\mathcal{G}_1$		Group $\mathcal{G}_2$	
Function	MSE	Function	MSE
$D(t)$	$1.942332e-06$	$C(t)$	$1.111295e-06$
$V_1(t)$	$2.428458e-04$	$V_2(t)$	$2.102767e-04$
$E(X_1(t))$	$3.193397e-05$	$E(X_2(t))$	$2.437206e-05$
$Var(X_1(t))$	$4.368352e-10$	$Var(X_2(t))$	$2.315756e-10$

We point out that the fit functions for Applications 1 and 2 are obtained by considering as data-generating process the model $X_1(t)$ and $X_2(t)$ in (10) and (12) for Application 1 and Eqs (11) and (12) for Application 2.

In the next section we consider the problem of testing if the influence over time of a therapy follows a given functional scheme and, in particular, if its application results in a constant modification of the natural parameters of the process.

4. Testing hypothesis about functions $C(t)$, $D(t)$ and $V(t)$

In the tumor context outlined in this paper, in order to evaluate the effectiveness of experimental therapies, the following questions are of special interest:

- Is the real effect on the growth rate of an anti-proliferative therapy null?
- Is the real effect on the death rate of a therapy that induces the death of cancer cells null?
- Does the therapy, or combination of therapies, affect the infinitesimal variance?
- Do the effects of a therapy or combination of therapies depend on time?
- Do the functions that model the effect of the therapy or combination of therapies have a specific form?

Since the functions included in model (2) represent different effects of a therapy, or combination of therapies on tumor growth, to answer these questions we propose to perform hypothesis testing about the functions $C(t)$, $D(t)$ and $V(t)$ in model (2) (note that models (10), (11) and (12) are particular cases of this).

The null hypothesis can be formulated in a unified way as

$$H_0 : H(t) = h(t),$$

with $H(t)$ any of the functions in model (2) and $h(t)$ a given function.

To test the null hypothesis we propose to use a bootstrap test (b-Test) based on the statistic $D = \sum_{j=0}^{n-1} |\widehat{H}(t_j) - h(t_j)|$. Calculation of values of this statistic is

based on a bootstrap procedure following the line proposed in Román-Román et al. [24, 30]. Concretely, the schema is the following:

- Generate m bootstrap samples of the considered model. Each bootstrap sample consists of d_k sample paths (depending of the treated group $\mathcal{G}_k$ being considered) simulated in the same way as the one previously exposed by taking function $h(t)$ in H_0 and the estimates of the parameters and the rest of functions via the procedure proposed in the previous section.
- Estimate $H(t)$ from the sample paths of each bootstrap sample, and calculate a value D_l, $l = 1, \ldots, m$, of the statistic D.
- Calculate the p-value as the proportion of values D_l greater than or equal to D.

The case $h(t) = h$ is of special interest because it means that the effect of therapy represented by $H(t)$ does not depend on time. Even more, $C(t) = 0$ means that the therapy has no anti-proliferative effect; $D(t) = 0$ signifies that the therapy does not induce the death of cancer cells, whereas $V(t) = 1$ leads to the non-influence of the therapy on the infinitesimal variance of the process.

In such case, the constant h to be included in H_0 has to be chosen. If it is not known a priori, as usual in applications, we propose to choose h as the value obtained from the ML estimation of $h(t) = h$ in model (2). Concretely,

- Testing $C(t)$ constant. In this case, $H_0 : C(t) = c$. The value of c is obtained from the ML estimate of the growth rate in model (2) by solving (18) in Appendix taking $C(t) = 0$, $D(t) = \widehat{D}(t)$ (if $D(t) \neq 0$), $V(t) = \widehat{V}(t)$ (if $V(t) \neq 1$), and considering $\beta = \widehat{\beta}$ and $\sigma = \widehat{\sigma}$, the ML estimates of β and σ for group $\mathcal{G}$. In this way we obtain $\widehat{\alpha - c}$, from which $c = \widehat{\alpha} - \widehat{\alpha - c}$.
- Testing $D(t)$ constant. Now, $H_0 : D(t) = d$, where the value d is obtained as in the previous case, by changing $C(t)$ with $D(t)$ and α with β. Note that in this case the equation (19) in Appendix must be solved.

- Testing $V(t)$ constant. This case is performed considering $H_0 : V(t) = v$, where the constant v is obtained from $v = \widehat{\sigma^2 v}/\widehat{\sigma^2}$, were $\widehat{\sigma^2 v}$ matches the ML estimate of σ^2 in the model (2) by solving (20) in Appendix taking $V(t) = 1$, $\alpha = \widehat{\alpha}$, $\beta = \widehat{\beta}$, $C(t) = \widehat{C}(t)$ (if $C(t) \neq 0$) and $D(t) = \widehat{D}(t)$ (if $D(t) \neq 0$).

4.1. Simulation study

In order to show the behavior of the proposed bootstrap tests we have performed a simulation study considering a control group $\mathcal{G}$ and two treated groups $\mathcal{G}_1$ and $\mathcal{G}_2$, modeled by (9), (10) and (12), respectively. The present study is limited to the case in which $\mathcal{G}_1$ is treated with an anti-proliferative therapy. The study in the case in which $\mathcal{G}_1$ is treated with a therapy that induces the death of cancer cells would be carried out in a similar way.

Once the models are estimated following the corresponding procedure in Section 3.1, we test hypotheses about all the functions included in (10) and (12) as follows.

1. For group $\mathcal{G}_1$, $H_0 : V_1(t) = v_1$ is tested, where v_1 is proposed following the comments mentioned above. Then, we test $H_0 : C(t) = c$, taking into account that:
   - If $H_0 : V_1(t) = v_1$ is not rejected, the value of c is determined considering $V_1(t) = v_1$.
   - If $H_0 : V_1(t) = v_1$ is rejected, c is determined considering $V_1(t) = \widehat{V}_1(t)$.
2. For group $\mathcal{G}_2$, $H_0 : V_2(t) = v_2$ is tested. To this end, v_2 is determined making use of $C(t) = c$, if $H_0 : C(t) = c$ is not rejected, or $C(t) = \widehat{C}(t)$ otherwise, and $D(t) = \widehat{D}(t)$. Then we test $H_0 : D(t) = d$, noting in this case that:
   - If $H_0 : V_2(t) = v_2$ is not rejected, d is determined making use of $C(t) = \widehat{C}(t)$, if $H_0 : C(t) = c$ is rejected, or $C(t) = c$ on the contrary, and $V_2(t) = v_2$.

- If $H_0 : V_2(t) = v_2$ is rejected, the value of d is determined considering $C(t) = \widehat{C}(t)$, if $H_0 : C(t) = c$ is rejected, or $C(t) = c$ otherwise, and $V_2(t) = \widehat{V}_2(t)$.

In the simulation study we have considered models (9), (10) and (12) with $\alpha = 0.5$, $\beta = 0.2$ and $\sigma = 0.01$ for all possible combinations of the functions in Table 3. Precisely, we consider 3 choices for functions $C(t)$, $D(t)$ and $V_1(t)$ and four cases for $V_2(t)$, obtaining 9 cases for group $\mathcal{G}_1$ and 108 for group $\mathcal{G}_2$. These functions have been selected in order to simulate the tumor growth in the groups treated with therapies of diverse effects, ranging from therapies that do not produce any improvement, until therapies that produce a significant reduction both in the mean relative volume of the tumor and in its variability.

For each model, 25 sample paths have been simulated over 51 equally time instants in $[0, 50]$. The estimates obtained for model (9), from the data of the control group $\mathcal{G}$, were $\widehat{\alpha} = 0.4972273$, $\widehat{\beta} = 0.1987757$ and $\widehat{\sigma} = 0.0100692$. Further, in each case for groups $\mathcal{G}_1$ and $\mathcal{G}_2$, the number of bootstrap samples used to perform each b-Test was $m = 1500$.

The complete simulation study is presented in schematic form in Supplementary Material. In each case, for groups $\mathcal{G}_1$ and $\mathcal{G}_2$, we show:

- the sample paths simulated for each model,
- the estimates of the functions included in each model,
- the results of the hypothesis tests listed above (concatenated b-Tests)
- the sample mean and variance functions,
- the theoretical mean and variance functions, together with their estimated versions, before and after the concatenated b-Tests are performed.

Table 3: Different values of functions $C(t)$, $V_1(t)$, $D(t)$ and $V_2(t)$ for the simulation study, being $\Lambda_1(t, \mu, \sigma^2)$ the density function of a lognormal distribution $\Lambda_1(\mu, \sigma^2)$.

$\mathbf{C(t)}$	$\mathbf{V_1(t)}$	$\mathbf{D(t)}$	$\mathbf{V_2(t)}$
0	1	0	1
0.025	0.49	-0.05	0.49
$0.005\,t$	$(0.7+10\,\Lambda_1(t,3,0.5))^2$	$\dfrac{-0.12\,t^2}{50+t(t-10)}$	$(0.7+10\,\Lambda_1(t,3,0.5))^2$
			$(0.7+15\,\Lambda_1(t,3,0.5))^2$

In this section we focus on two cases that have a specific meaning. These two cases show how the proposed tests allow obtaining a better estimation of the mean functions and variances of the simulated processes.

- **Case 1**. $C(t) = 0$, $V_1(t) = 1$, $D(t) = 0$ and $V_2(t) = 1$.

  In this case, in the group $\mathcal{G}_1$, the anti-proliferative effect of the first therapy is null, and this therapy also does not affect the infinitesimal variance of process $X_1(t)$. Moreover, in group $\mathcal{G}_2$, the effect of therapy that induces the death of cancer cells is null, and the combined effect of the two therapies does not affect the infinitesimal variance of the process $X_2(t)$.

  Figure 3 shows the simulated sample paths of models (9), (10) and (12).

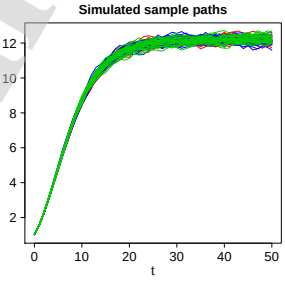


Figure 3: Simulated sample paths of models (9), (10) and (12), in red, blue and green, respectively, with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0$, $V_1(t) = 1$, $D(t) = 0$ and $V_2(t) = 1$.

Model (10) has been adjusted from the data of treated group $\mathcal{G}_1$ with

$\alpha = \widehat{\alpha}$, $\beta = \widehat{\beta}$ and $\sigma = \widehat{\sigma}$, obtaining the estimates $\widehat{C}(t)$ and $\widehat{V}_1(t)$. These estimates are shown in Figure 4 as well as the estimated mean and variance functions of the process $X_1(t)$. Specifically, in Figures 4(a) and 4(b), the red points correspond to the estimated values $\widehat{C}_i$ and $\widehat{V}_{1,i}$, the red solid lines represent the estimated functions $\widehat{C}(t)$ and $\widehat{V}_1(t)$, while the black solid lines indicate the theoretical functions. In Figures 4(c) and 4(d), the red points correspond to the sample mean and variance functions, respectively; the red solid lines represent the estimated mean and variance functions of the process $X_1(t)$ whereas the black solid lines indicate the theoretical mean and variance functions of process $X_1(t)$. In all the cases, the absolute difference functions between the simulated and fitted function are also represented in green.

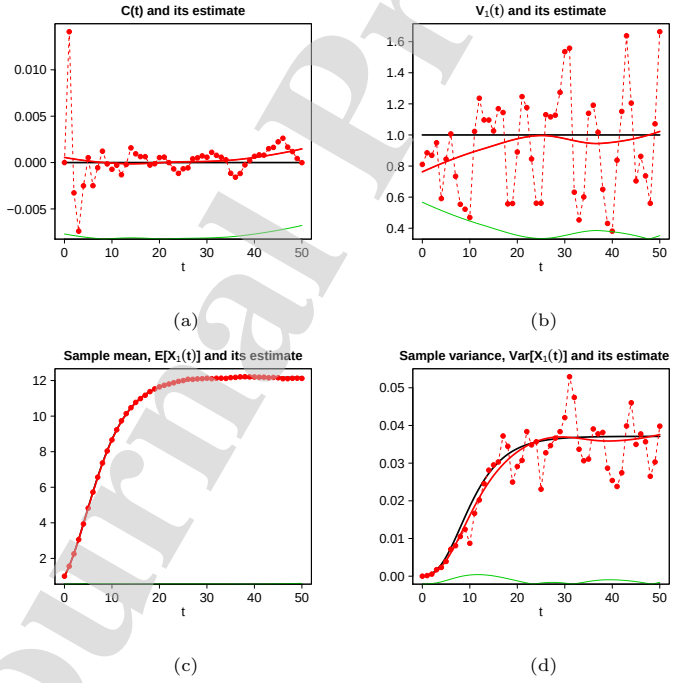


Figure 4: Fit of simulated data of model (10) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0$ and $V_1(t) = 1$. The absolute difference functions between the simulated and fitted function are represented in green.

Figure 4 seems to indicate that the estimated values $\widehat{C}_i$ and $\widehat{V}_{1,i}$ vary around values close to 0 and 1, respectively, and it makes sense to test if the functions $C(t)$ and $V_1(t)$ are constant.

First, we test if $V_1(t)$ is constant. The constant is chosen via the ML estimation as previously indicated. In this case, $v_1 = 0.9710093$. The value of the D-statistics is $D = 2.5203844$ and the associated p-value is 0.974, so there is no evidence to reject that the effect of the anti-proliferative therapy on the infinitesimal variance of the process that models tumor growth does not depend on time. Moreover, $v_1 \approx 1$ suggests that the anti-proliferative therapy hardly affects such infinitesimal variance.

Then, under the assumption $V_1(t) = 0.9710093$, we test if $C(t)$ is constant, i.e. $H_0 : C(t) = c$ where $c = 0.0002401$ has been determined as described before. The value of the D-statistics is $D = 0.0158729$ and the associated p-value is 0.755, so there is no evidence to reject that the effect of the anti-proliferative therapy on the rate of growth does not depend on time. In fact, since $c \approx 0$, we can conclude that the supposed anti-proliferative effect of the therapy on the growth rate has been almost null.

Figure 5, similarly to Figures 4(c) and 4(d), shows the estimated mean and variance functions in the group $\mathcal{G}_1$ with $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\sigma}$, $\widehat{V}_1(t) = 0.9710093$ and $\widehat{C}(t) = 0.0002401$, together with the sample and theoretical mean and variance functions of process $X_1(t)$. Observe how now the estimated mean and variance functions better reproduce the theoretical ones.

Next, from data of the treated group $\mathcal{G}_2$, model (12) have been adjusted by using $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\sigma}$ and $\widehat{C}(t) = 0.0002401$. As in Figure 4, the estimated functions $\widehat{D}(t)$ and $\widehat{V}_2(t)$ in Figure 6 are plotted as well as the estimated mean and variance functions of the process $X_2(t)$, showing how the values $\widehat{D}_i$ and $\widehat{V}_{2,i}$ are close to 0 and 1, respectively, which leads us to test whether the functions $D(t)$ and $V_2(t)$ are constant.

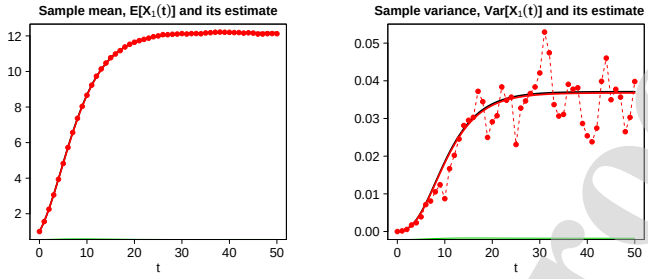


Figure 5: Fit of simulated data of model (10) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0$ and $V_1(t) = 1$, assuming that it is accepted first that $V_1(t)$ is constant, and then, that $C(t)$ is constant. The absolute difference functions between the simulated and fitted function are represented in green.

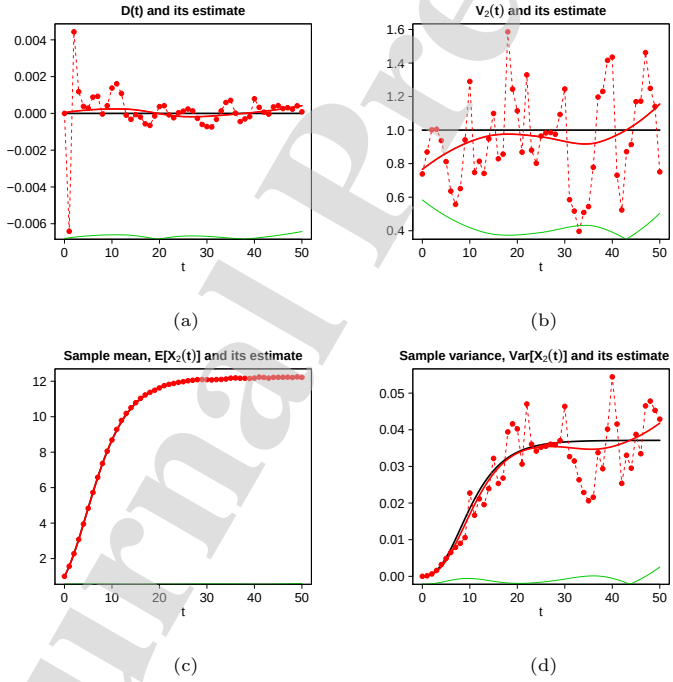


Figure 6: Fit of simulated data of model (12) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0$, $D(t) = 0$ and $V_2(t) = 1$. The absolute difference functions between the simulated and fitted function are represented in green.

First, we test the hypothesis $H_0 : V_2(t) = v_2$, where now $v_2 = 0.9778234$.

The bootstrap test results in $D = 2.9449791$ and the associated p-value= 0.931 and, therefore, there is no evidence to reject that the effect of the combination of therapies on the infinitesimal variance of the process $X_2(t)$ does not depend on the time. In addition, as $v_2 \approx 1$, the combination of therapies has a negligible effect on such infinitesimal variance.

Thus, assuming $V_2(t) = 0.9778234$, we test $H_0 : D(t) = d$, being now $d = 0.0002078$. The value of the D-statistics is 0.0084735 and the p-value is 0.897, so that there is no evidence to reject that the effect of the therapy inducing the death of cancer cells does not depend on time. Furthermore, since $d \approx 0$, we can conclude that the therapy has hardly induced the death of cancer cells.

Figure 7 shows the estimated mean and variance functions in the group $\mathcal{G}_2$ for $\widehat{\alpha} = 0.4972273$, $\widehat{\beta} = 0.1987757$, $\widehat{\sigma} = 0.0100692$, $\widehat{C}(t) = 0.0002401$, $\widehat{V}_2(t) = 0.9778234$ and $\widehat{D}(t) = 0.0002078$, together with the sample and the theoretical mean and variance functions of process $X_2(t)$. Comparing this figure with Figures 6(c) and 6(d), we can see that the estimated mean and variance functions better reproduce the theoretical ones.

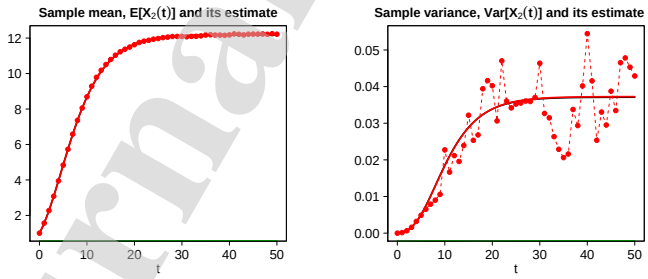


Figure 7: Fit of simulated data of model (12) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0$, $D(t) = 0$ and $V_2(t) = 1$, knowing that $C(t)$ is constant and assuming that it is accepted at first that $V_2(t)$ is constant, and then, that $D(t)$ is constant

In Figure 8 the Gaussian kernel density estimations of the D-statistics for the tests just discussed are plotted based on $m = 1500$ runs. The

bandwidth is chosen by using pilot estimation of derivatives as indicated in [31]. To summarize the results of the tests the values of the D-statistics (green points) and the critical region with significance 0.05 (red) are also shown.

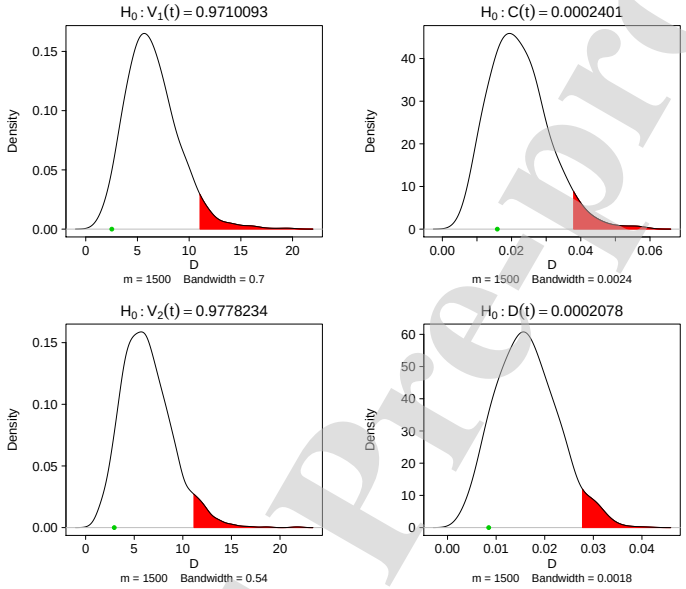


Figure 8: Gaussian kernel density estimation of D-statistics for the tests associated to Case 1. Green points are the values of the D-statistics in our simulation experiment. In red the critical region with significance 0.05 is shown.

- **Case 2.** $C(t) = 0.025$, $V_1(t) = 0.49$, $D(t) = -0.05$ and $V_2(t) = 0.49$

  In this case, in group $\mathcal{G}_1$, an anti-proliferative therapy that affects the infinitesimal variance of the process $X_1(t)$ is considered, although its effects do not depend on time. In group $\mathcal{G}_2$, the effect of the therapy that induces the death of cancer cells does not depend on time and such therapy does not affect the infinitesimal variance previously modified by the anti-proliferative therapy.

  Figure 9 shows the simulated sample paths of models (9), (10) and (12) in red, blue and green, respectively.

Figure 10 shows the estimated functions $\widehat{C}(t)$ and $\widehat{V}_1(t)$ in model (10) as well as the estimated mean and variance functions of process $X_1(t)$. We can see that the estimated values $\widehat{C}_i$ and $\widehat{V}_{1,i}$ vary around values close to 0.025 and 0.49, respectively, which suggests that the functions $C(t)$ and $V_1(t)$ could be constant.

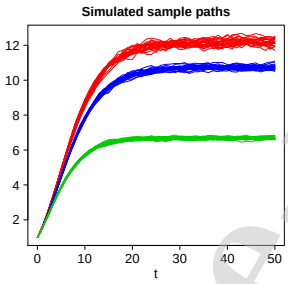


Figure 9: Simulated sample paths of models (9), (10) and (12), in red, blue and green, respectively, with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0.025$, $V_1(t) = 0.49$, $D(t) = -0.05$ and $V_2(t) = 0.49$.

First we test the hypothesis $H_0 : V_1(t) = v_1$ where $v_1 = 0.4761118$. The value of the D-statistics is $D = 1.0633344$ and the p-value is 0.996, so there is no evidence to reject that the effect of the anti-proliferative therapy on the infinitesimal variance of the process $X_1(t)$ does not depend on time.

Then, under the assumption $V_1(t) = 0.4761118$, we test $H_0 : C(t) = c$ with $c = 0.0248451$. The bootstrap test provides $D = 0.0234247$ and a p-value of 0.334, so there is no evidence to reject that the effect of the anti-proliferative therapy on the growth rate does not depend on time.

Figure 11 shows the estimated mean and variance functions in the group $\mathcal{G}_1$ with $\widehat{\alpha} = 0.4972273$, $\widehat{\beta} = 0.1987757$, $\widehat{\sigma} = 0.0100692$, $\widehat{C}(t) = 0.0248451$ and $\widehat{V}_1(t) = 0.4761118$, together with the sample and the theoretical mean and variance functions of the process $X_1(t)$. It is clear that now the estimated mean and variance functions reproduce more appropriately the theoretical ones.

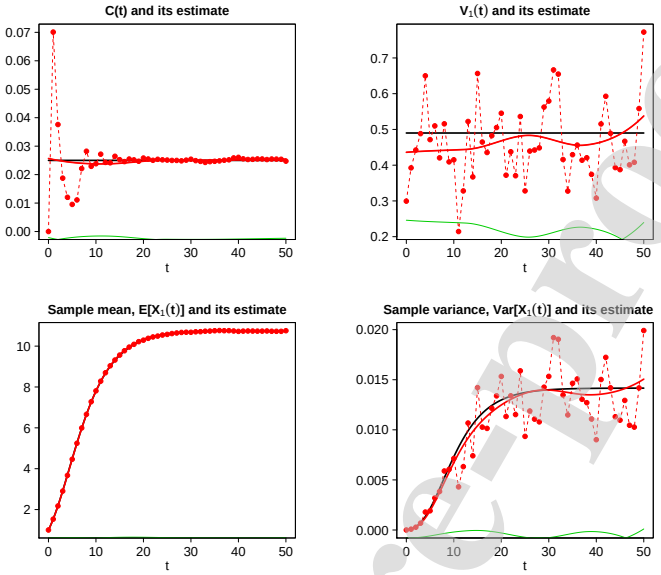


Figure 10: Fit of simulated data of model (10) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0.025$ and $V_1(t) = 0.49$. The absolute difference functions between the simulated and fitted function are represented in green.

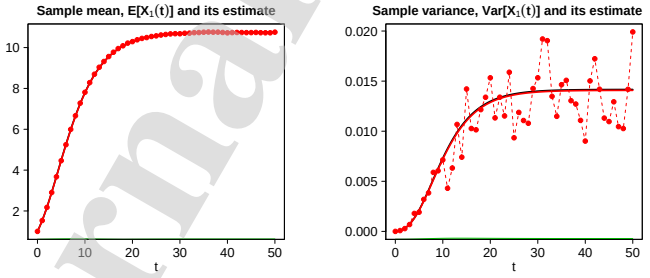


Figure 11: Fit of simulated data of model (10) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0.025$ and $V_1(t) = 0.49$, assuming that it is accepted first that $V_1(t)$ is constant, and then, that $C(t)$ is constant. The absolute difference functions between the simulated and fitted function are represented in green.

Figure 12 shows the estimated functions $\widehat{D}(t)$ and $\widehat{V}_2(t)$ in model (12) by using $\widehat{\alpha}$, $\widehat{\beta}$, $\widehat{\sigma}$ and $\widehat{C}(t) = 0.0248451$, as well as the estimated mean and

variance functions of process $X_2(t)$. The estimated values $\widehat{D}_i$ and $\widehat{V}_{2,i}$ in this figure vary around values close to -0.05 and 0.49, respectively, and it seems reasonable to test whether the functions $D(t)$ and $V_2(t)$ are constant.

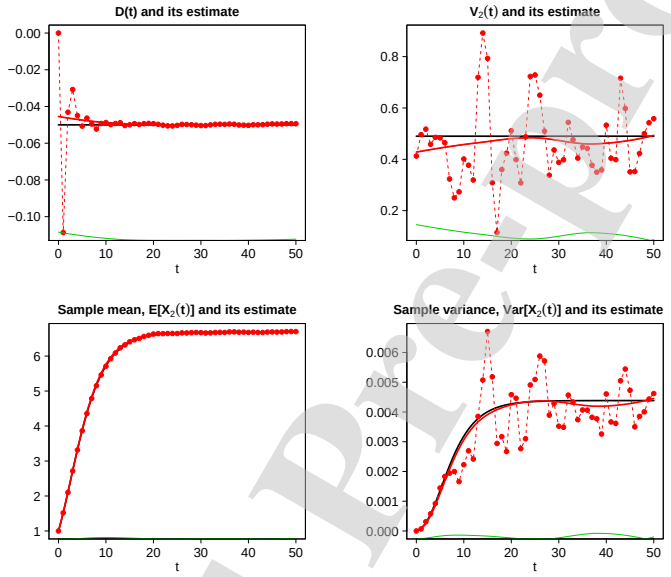


Figure 12: Fit of simulated data of model (12) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0.025$, $D(t) = -0.05$ and $V_2(t) = 0.49$.

First we test $H_0 : V_2(t) = v_2$ where $v_2 = 0.4842422$. The value of the D-statistics is $D = 0.9347132$ and the p-value is 0.981, so that there is no evidence to reject that the effect of the combination of therapies on the infinitesimal variance of the process $X_2(t)$ is independent on time.

Hence, under the assumption $V_2(t) = 0.4842422$, we test $H_0 : D(t) = d$ where the proposed value for d is -0.0497542. The bootstrap test results in $D = 0.0359978$ and p-value=0.321 and consequently there is no evidence to reject that the therapy that induces the death of cancer cells does not depend on time.

Figure 13 shows the estimated mean and variance functions in group $\mathcal{G}_2$

with $\widehat{\alpha} = 0.4972273$, $\widehat{\beta} = 0.1987757$, $\widehat{\sigma} = 0.0100692$, $\widehat{C}(t) = 0.0248451$, $\widehat{V}_2(t) = 0.4842422$ and $\widehat{D}(t) = -0.0497542$, together with the theoretical and the sample mean and variance functions of the process $X_2(t)$. Again, the estimated mean and variance functions better reproduce the theoretical ones.

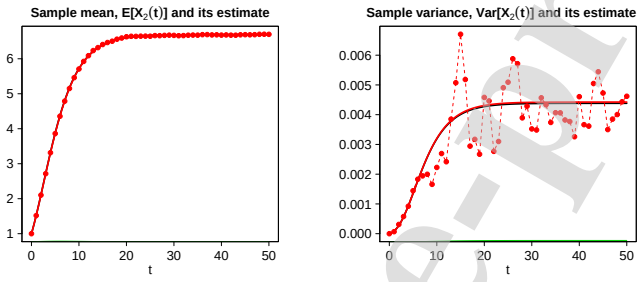


Figure 13: Fit of simulated data of model (12) with $\alpha = 0.5$, $\beta = 0.2$, $\sigma = 0.01$, $C(t) = 0.025$, $D(t) = -0.05$ and $V_2(t) = 0.49$ knowing that $C(t)$ is constant and assuming that it is accepted first that $V_2(t)$ is constant, and then that $D(t)$ is constant.

As in Figure 8, the Gaussian kernel density estimations of the D-statistics for the tests just discussed are plotted in Figure 14.

5. Application to real data of tumor growth

In this section we apply the stochastic process introduced in this paper to model experimental data, obtained in mice, in order to study the effect of two treatments on ovarian cancer.

In particular, we analyze the effects of Carboplatin and Paclitaxel treatments on the growth of OVA014HENp9 tumor from data of three experimental groups of 9, 8 and 8 mice. These data has been provided by the Laboratory of Preclinical Investigation (LIP) that belongs to the Translational Research Department of the Institute Curie, Paris. Carboplatin plus paclitaxel regimen remains the standard chemotherapy for the initial treatment of ovarian cancer and it is less toxic and easier to administrate compared to other drug combinations (Ozols et al. [32]).

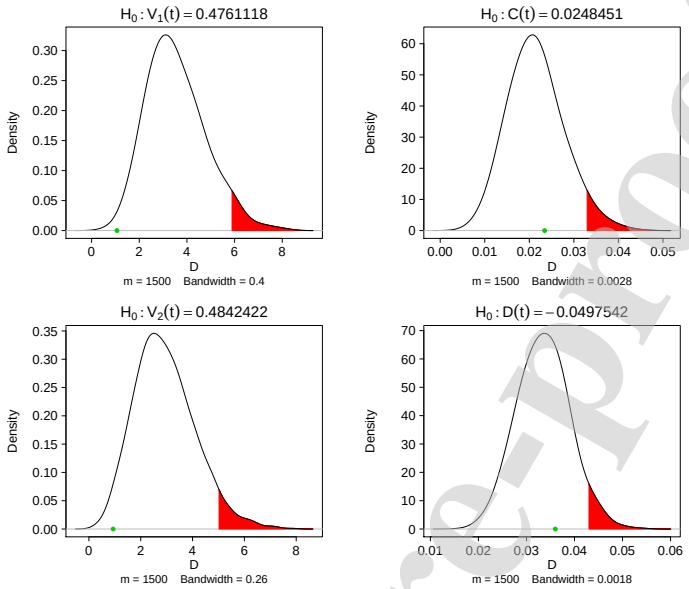


Figure 14: Gaussian kernel density estimation of D-statistics for the tests associated to Case 2. Green points are the values of the D-statistics in our simulation experiment. In red the critical region with significance 0.05 is shown.

The first group, $\mathcal{G}$, was a control (untreated); the second group, $\mathcal{G}_1$, was treated with Carboplatin (66mg/kg/day the days 1 and 22); and the third group, $\mathcal{G}_2$, received Carboplatin(idem)+Paclitaxel (12mg/kg/week over a period of six weeks). The relative volume of tumor was measured at days 1, 4, 11, 16, 19, 31, 34, 38, 41, 53, 59 and 66.

Figures 15 and 16 show the sample paths and the sample mean and variance, respectively, of the relative tumor volume for the three experimental groups as a function of the days after starting the treatment.

The ML estimation of the parameters in control group provide $\widehat{\alpha} = 0.06964254$, $\widehat{\beta} = 0.01238329$, $\widehat{\sigma} = 0.08964128$.

Since the therapy with Carboplatin induces the death of cancer cells, we have adjusted the model (11) to the data of treated group $\mathcal{G}_1$. Figure 17 shows the estimates of the $D(t)$ and $V_1(t)$ functions as well as the fit of the sample

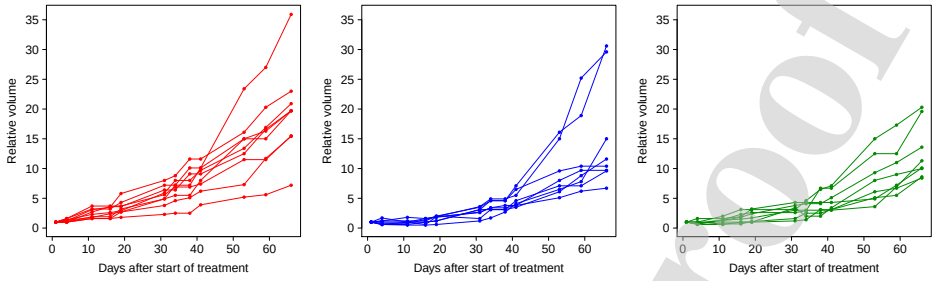


Figure 15: Sample paths of relative volume of tumor in control group (a), and Carboplatin (b) and Carboplatin+Paclitaxel (c) treated groups.

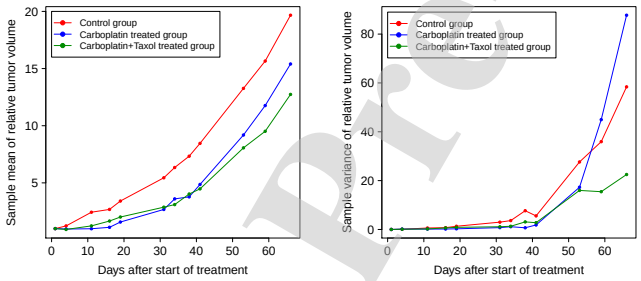


Figure 16: Sample mean and variance of the relative tumor volume in control and treated groups.

means and variances of data by using $E(\widehat{X}_1(t))$ and $Var(\widehat{X}_1(t))$, respectively.

In the same way, since the therapy with Paclitaxel is anti-proliferative, we have adjusted the model (12) to the data of the treated group $\mathcal{G}_2$. Figure 18 shows the estimates of the $C(t)$ and $V_2(t)$ functions as well as the fit of the sample means and variances of data by using $E(\widehat{X}_2(t))$ and $Var(\widehat{X}_2(t))$, respectively.

The results of the fitting function $D(t)$ (Figure 17) show that the Carboplatin treatment is effective in the first 15-20 days in which it present a negative peak, then it becomes ineffective. The infinitesimal variance $V_1(t)$ seems to be greatly influenced by the therapy when it is effective, after that the variability of the process restore to natural constant values. Concerning the treated Carboplatin+Paclitaxel group $\mathcal{G}_2$ (Figure 18), we observe that the therapy is effective

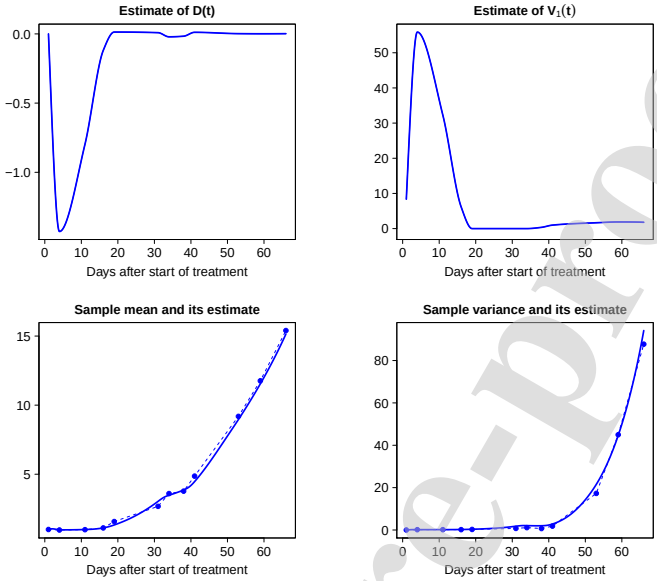


Figure 17: Fit of model (11) in Carboplatin treated group (estimates in solid line).

in the first days of the treatment, then its effectivenes declines corresponding to a negative bump, followed from values close to zero. The function $V_2(t)$ presents a similar behaviour with respect to $V_1(t)$, although it shows a lower peak.

The estimated models in both treated groups provide a good fit of the sample means and variances of the relative volume of tumor. Table 4 presents the mean squared errors between the sample mean and variance functions of the simulated process and the estimated ones, that is

$$MeanMSE = \frac{1}{n}\sum_{j=1}^{n}(m_j - \hat{m}_j)^2, \qquad VarMSE = \frac{1}{n}\sum_{j=1}^{n}(\sigma_j^2 - \hat{\sigma}_j^2)^2$$

where (m_j, σ_j^2) are the values of the sample mean and variance functions at t_j, $j = 1, \ldots, n$ whereas $(\hat{m}_j, \hat{\sigma}_j^2)$ are the estimated ones.

Taking into account the estimates of functions $D(t)$, $V_1(t)$, $C(t)$ and $V_2(t)$, it seems reasonable to test only if $C(t)$ is constant. However, we have tested, in a concatenated form, if each of the functions could be constant, as in the simulation study in Section 4.1. The corresponding constants to be included in

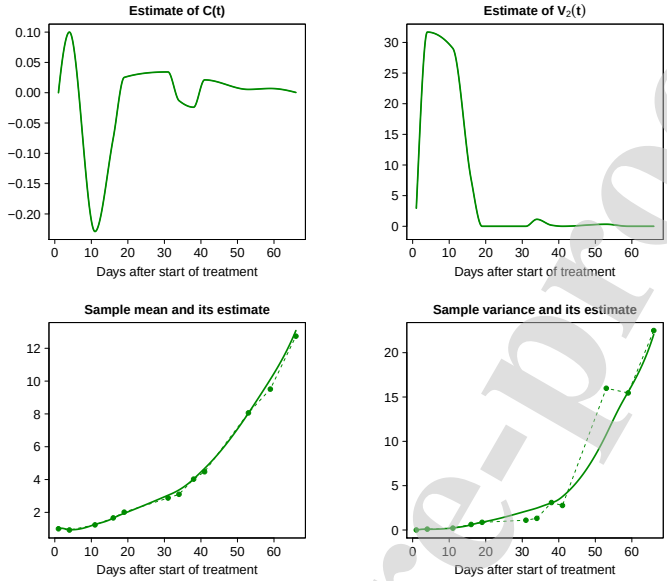


Figure 18: Fit of model (12) in Carboplatin+Paclitaxel treated group (estimates in solid line).

Table 4: Mean squared errors for the fits of the sample means and variances in the treated groups.

MSEs	Carboplatin	Carboplatin+Paclitaxel
MeanMSEs	0.0445	0.0595
VarMSEs	5.3507	2.5464

the null hypothesis are estimated by ML as described in Section 4.

For group $\mathcal{G}_1$, $H_0 : V_1(t) = 9.612252$ was tested first. The associated p-value was 0.04 and therefore we reject that $V_1(t)$ be constant. Then, we have tested $H_0 : D(t) = d$, where $d = 0.001296$ is determined by ML considering $V_1(t) = \widehat{V}_1(t)$. The test produced a p-value of 0.03 and we also reject that $D(t)$ be constant.

Next, for group $\mathcal{G}_2$, $H_0 : V_2(t) = 8.097848$ was tested (making use of $D(t) = \widehat{D}(t)$). This hypothesis is rejected with a p-value of 0.01. Finally, we have tested $H_0 : C(t) = c$, where $c = 0.023195$ is determined by ML considering

$V_2(t) = \widehat{V}_2(t)$. The resulting p-value was 0.03 and we must also reject that $C(t)$ be constant.

Thus, we can conclude that, in group $\mathcal{G}_1$, the effect of Carboplatin on the death of cancer cells and the infinitesimal variability of relative volume of tumor is time-dependent. In a similar way, in group $\mathcal{G}_2$, the same comment can be done about the combined effect of Carboplatin and Paclitaxel on the growth and death of cancer cells, as well as on the infinitesimal variability of relative volume of tumor. Therefore, based on the tests carried out, we can conclude that the therapies applied are time dependent, so the models (11) and (12) seem to be appropriated to describe the combined effect of the two therapies.

Appendix A. Maximum likelihood estimates of the parameters of the process

The objective of this appendix is to provide the ML estimation of the parameters of the process in model (2) for known $C(t)$, $D(t)$ and $V(t)$ functions. In addition we provide the ML estimation of each parameter for known values of the rest ones, which will be useful when establishing the null hypotheses $H_0 : H(t) = h$, with $H(t)$ any of the functions in model (2).

From (5) in Section 2.1, the transition pdf of the process can be obtained, resulting in

$$X(t)|X(s) = y \sim \Lambda_1 \left[\bar{k}(t|s) \ln y + \theta(t|s), \sigma^2 \Omega(t|s)\right] \tag{16}$$

where

$$\theta(t|\tau) = \int_\tau^t \left(\alpha - C(s) - \frac{\sigma^2}{2} V(s)\right) \bar{k}(t|s)\, ds,$$
$$\Omega(t|\tau) = \int_\tau^t \bar{k}^2(t|s) V(s)\, ds$$

Let us consider a discrete sampling $\{x_{ij},\ i = 1, \ldots, d;\ j = 0, \ldots, n_i - 1\}$ of the process based on d sample paths at times t_{ij}, $(i = 1, \ldots, d,\ j = 0, \ldots, n_i - 1)$ with $t_{i0} = t_0$ and $x_{i0} = x_0$, $i = 1, \ldots, d$. Denote by $\mathbf{X} = \left(\mathbf{X}_1^T | \cdots | \mathbf{X}_d^T\right)^T$, where $\mathbf{X}_i = (X_{i0}, \ldots, X_{i,n_i-1})^T$, $i = 1, \ldots, d$, with $X_{ij} = X(t_{ij})$, $j = 0, \ldots, n_i - 1$.

By taking $X(t_0)$ a degenerate random variable, i.e. $P[X(t_0) = x_0] = 1$, from (16), the probability density function of $\mathbf{X}$ is

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{d} \prod_{j=1}^{n_i-1} \frac{\exp\left(-\frac{\left[\delta_\beta^{ij} - \theta_{\boldsymbol{\xi}}^{ij}\right]^2}{2\sigma^2\Omega_\beta^{ij}}\right)}{x_{ij}\sigma\sqrt{2\pi\Omega_\beta^{ij}}}$$

where $\delta_\beta^{ij} = \ln x_{ij} - \bar{k}_\beta^{ij} \ln x_{i,j-1}$, $\theta_{\boldsymbol{\xi}}^{ij} = \theta(t_{ij}|t_{i,j-1})$ and $\Omega_\beta^{ij} = \Omega(t_{ij}|t_{i,j-1})$, with $\bar{k}_\beta^{ij} = \bar{k}(t_{ij}|t_{i,j-1}) = exp\left(-\int_{t_{i,j-1}}^{t_{ij}} (\beta - D(s))ds\right)$, $i = 1, 2, \ldots, d$, $j = 1, \ldots, n_i - 1$, and $\boldsymbol{\xi} = (\alpha, \beta, \sigma^2)^T$.

Then, for a fixed value $\mathbf{x}$ of the sample and known $C(t)$, $D(t)$ and $V(t)$ functions, the log-likelihood function is

$$L_{\mathbf{x}}(\boldsymbol{\xi}) = -\frac{n\ln(2\pi)}{2} - \frac{n\ln\sigma^2}{2} - \frac{Z_\beta + \Phi_{\boldsymbol{\xi}} - 2\Gamma_\xi}{2\sigma^2} - \frac{1}{2}\Upsilon_\beta - \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} ln\, x_{ij} \quad (17)$$

where $n = \sum_{i=1}^{d}(n_i - 1)$, $Z_\beta = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{(\delta_\beta^{ij})^2}{\Omega_\beta^{ij}}$, $\Phi_{\boldsymbol{\xi}} = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{(\theta_\xi^{ij})^2}{\Omega_\beta^{ij}}$, $\Gamma_{\boldsymbol{\xi}} = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} \frac{\delta_\beta^{ij}\theta_\xi^{ij}}{\Omega_\beta^{ij}}$ and $\Upsilon_\beta = \sum_{i=1}^{d}\sum_{j=1}^{n_i-1} ln\,\Omega_\beta^{ij}$.

In order to obtain the ML estimate of α, β and σ^2 we denote

$$\Psi_{\beta,ij}^{l,m,p,q} = \int_{t_{i,j-1}}^{t_{ij}} (t_{ij} - s)^l \left(C(s)\right)^m \left(V(s)\right)^p \left(\bar{k}(t_{ij}|s)\right)^q ds,$$

from which we deduce:

$$\frac{\partial \Psi_{\beta,ij}^{l,m,p,q}}{\partial \beta} = -q\Psi_{\beta,ij}^{l+1,m,p,q}$$

$$\theta_{\boldsymbol{\xi}}^{ij} = \alpha\Psi_{\beta,ij}^{0,0,0,1} - \Psi_{\beta,ij}^{0,1,0,1} - \frac{\sigma^2}{2}\Psi_{\beta,ij}^{0,0,1,1}$$

$$\Omega_\beta^{ij} = \Psi_{\beta,ij}^{0,0,1,2}$$

The likelihood equations are:

$$\frac{\partial L_{\mathbf{x}}}{\partial \alpha} = -\frac{1}{2\sigma^2}\left(\frac{\partial \Phi_{\boldsymbol{\xi}}}{\partial \alpha} - 2\frac{\partial \Gamma_{\boldsymbol{\xi}}}{\partial \alpha}\right) = 0$$

$$\frac{\partial L_{\mathbf{x}}}{\partial \beta} = -\frac{1}{2\sigma^2}\left(\frac{\partial Z_\beta}{\partial \beta} + \frac{\partial \Phi_{\boldsymbol{\xi}}}{\partial \beta} - 2\frac{\partial \Gamma_{\boldsymbol{\xi}}}{\partial \beta}\right) + \frac{1}{2}\frac{\partial \Upsilon_\beta}{\partial \beta} = 0$$

$$\frac{\partial L_{\mathbf{x}}}{\partial \sigma^2} = -\frac{1}{2\sigma^2}\left(n - \frac{1}{\sigma^2}(Z_\beta + \Phi_{\boldsymbol{\xi}} - 2\Gamma_{\boldsymbol{\xi}}) + \frac{\partial \Phi_{\boldsymbol{\xi}}}{\partial \sigma^2} - 2\frac{\partial \Gamma_{\boldsymbol{\xi}}}{\partial \sigma^2}\right) = 0$$

or equivalently, after calculus,

$$2\alpha\, X_1^\beta - 2X_2^\beta - \sigma^2 X_3^\beta - 2X_4^\beta = 0 \tag{18}$$

$$X_5^\beta - X_6^\beta - X_7^\beta - X_8^\beta - X_9^\beta + 2X_{10}^\beta + X_{11}^\beta + \frac{\sigma^2}{2} X_{12}^\beta = 0 \tag{19}$$

$$\frac{\sigma^4}{4} X_{13}^\beta + n\sigma^2 - (Z_\beta + \alpha^2\, X_1^\beta - 2\alpha\, X_2^\beta - 2\alpha\, X_4^\beta + X_{14}^\beta + 2X_{15}^\beta) = 0 \tag{20}$$

where

$$X_1^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\left(\Psi_{\beta,ij}^{0,0,0,1}\right)^2}{\Psi_{\beta,ij}^{0,0,1,2}}, \qquad X_2^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\Psi_{\beta,ij}^{0,1,0,1} \Psi_{\beta,ij}^{0,0,0,1}}{\Psi_{\beta,ij}^{0,0,1,2}},$$

$$X_3^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\Psi_{\beta,ij}^{0,0,1,1} \Psi_{\beta,ij}^{0,0,0,1}}{\Psi_{\beta,ij}^{0,0,1,2}}, \quad X_4^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\delta_\beta^{ij} \Psi_{\beta,ij}^{0,0,0,1}}{\Psi_{\beta,ij}^{0,0,1,2}},$$

$$X_5^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\bar{k}_\beta^{ij} \Delta_{ij} \theta_{\boldsymbol{\xi}}^{ij} \ln x_{ij}}{\Psi_{\beta,ij}^{0,0,1,2}}, \quad X_6^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\theta_{\boldsymbol{\xi}}^{ij} \varphi_{\boldsymbol{\xi}}^{ij}}{\Psi_{\beta,ij}^{0,0,1,2}},$$

$$X_7^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\left(\theta_{\boldsymbol{\xi}}^{ij}\right)^2 \Psi_{\beta,ij}^{1,0,1,2}}{\left(\Psi_{\beta,ij}^{0,0,1,2}\right)^2}, \quad X_8^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\left(\delta_\beta^{ij}\right)^2 \Psi_{\beta,ij}^{1,0,1,2}}{\left(\Psi_{\beta,ij}^{0,0,1,2}\right)^2},$$

$$X_9^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\delta_\beta^{ij} \bar{k}_\beta^{ij} \Delta_{ij} \ln x_{ij}}{\Psi_{\beta,ij}^{0,0,1,2}}, \quad X_{10}^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\delta_\beta^{ij} \Psi_{\beta,ij}^{1,0,1,2} \theta_{\boldsymbol{\xi}}^{ij}}{\left(\Psi_{\beta,ij}^{0,0,1,2}\right)^2},$$

$$X_{11}^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\delta_\beta^{ij} \varphi_{\boldsymbol{\xi}}^{ij}}{\Psi_{\beta,ij}^{0,0,1,2}}, \qquad X_{12}^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\Psi_{\beta,ij}^{1,0,1,2}}{\Psi_{\beta,ij}^{0,0,1,2}},$$

$$X_{13}^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\left(\Psi_{\beta,ij}^{0,0,1,1}\right)^2}{\Psi_{\beta,ij}^{0,0,1,2}}, \qquad X_{14}^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\left(\Psi_{\beta,ij}^{0,1,0,1}\right)^2}{\Psi_{\beta,ij}^{0,0,1,2}},$$

$$X_{15}^\beta = \sum_{i=1}^{d} \sum_{j=1}^{n_i-1} \frac{\delta_\beta^{ij} \Psi_{\beta,ij}^{0,1,0,1}}{\Psi_{\beta,ij}^{0,0,1,2}},$$

being $\Delta_{ij} = t_{i,j+1} - t_{ij}$ and $\varphi_{\boldsymbol{\xi}}^{ij} = -\alpha \Psi_{\beta,ij}^{1,0,0,1} + \Psi_{\beta,ij}^{1,1,0,1} + \frac{\sigma^2}{2} \Psi_{\beta,ij}^{1,0,1,1}$.

6. Conclusions

Mathematical modeling of tumor growth can help investigators to improve the design of preclinical or clinical trials and to better predict treatment outcome. Actually a comprehensive description of tumor dynamics during therapy in preclinical setting allows to accurately compare different schemes of drug administration. Ultimately, preclinical correlations between tumor growth dynamics during treatment and the efficacy of drug(s) could help to tailor the schedule to be proposed to patients in clinical trials.

In this paper a modified Gompertz diffusion process including exogenous factors in its infinitesimal moments has been considered in order to model both the effect of anti-proliferative and/or cell death-induced therapies. A procedure to estimate the parameters and time functions included in the model has been proposed, and our simulation studies show how said procedure adequately reproduces both the parameters and the form of the functions involved, as well as the mean and variance of the simulated data. In addition, from the estimated model, we have provided bootstrap tests about the form of the true functions in the model. The estimated functions $C(t)$, $D(t)$ and $V(t)$ that finally result allow to understand how therapies affect tumor growth. Thus, our model could constitute a valuable tool to adjust the drug administration scheme in the preclinical setting, in order to improve the efficacy of treatment, and to optimize the schedule to be proposed to patients in clinical trials.

Acknowledgments

The authors are very grateful to Dr. Didier Decaudin (Institute Curie, Paris) for providing the data used in this research. They acknowledge the constructive criticism of anonymous reviewers on the earlier version of this paper.

This work was supported in part by the Ministerio de Economía, Industria y Competitividad, Spain, under Grant MTM2017-85568-P and by the Consejería de Economía y Conocimiento de la Junta de Andalucía, Spain under Grant A-FQM-456-UGR18 and by MIUR - PRIN 2017, Project “Stochastic Models

for Complex Systems" no. 2017JFFHSH. The authors Albano and Giorno are members of the research group GNCS-INdAM.

References

[1] J. M. Steele, Stochastic Calculus and Financial Applications, Vol. 45 of Stochastic Modelling and Applied Probability, Springer Science and Business Media, 2012.

[2] S. E. Shreve, Stochastic Calculus for Finance II: Continuous-Time Models, Springer Science and Business Media, 2004.

[3] L. J. S. Allen, An Introduction to Stochastic Processes with Applications to Biology, 2nd Edition, CRC Press, 2010.

[4] D. J. Wilkinson, Stochastic modelling for quantitative description of heterogeneous biological systems, Nature Reviews Genetics 10 (2009) 122–133. doi:10.1038/nrg2509.

[5] H. Ahn, H. Moon, S. Kim, K. R. L., A newton-based approach for attributing tumor lethality in animal carcinogenicity studies, Computational Statistics and Data Analysis 38 (2002) 263–283. doi:10.1016/S0167-9473(01)00041-X.

[6] A. Ochab-Marcinek, E. Gudowska-Nowak, Population growth and control in stochastic models of cancer development, Physica A: Statistical Mechanics and its Applications 343 (15) (2004) 557–572. doi:10.1016/j.physa.2004.06.071.

[7] A. Talkington, R. Durrett, Estimating tumor growth rates in vivo, Bulletin of Mathematical Biology 77 (2015) 1934–1954. doi:10.1007/s11538-015-0110-8.

[8] H. P. de Vladar, J. A. Gonzales, Dynamic response of cancer under the influence of immunological activity and therapy, Journal of Theoretical Biology 227 (2004) 335–348. doi:10.1016/j.jtbi.2003.11.012.

[9] L. Ferrante, S. Bompadre, L. Leone, M. P. Montanari, A stochastic formulation of the gompertzian growth model for in vitro bactericidal kinetics: parameter estimation and extinction probability, Biometrical Journal 47 (3) (2005) 309–318. doi:10.1002/bimj.200410125.

[10] C. F. Lo, Stochastic gompertz model of tumor cell growth, Journal of Theoretical Biology 248 (2) (2007) 317–321. doi:10.1016/j.jtbi.2007.04.024.

[11] C. F. Lo, A modified stochastic gompertz model for tumor cell growth, Computational and Mathematical Methods in Medicine 11 (1) (2008) 3–11. doi:10.1080/17486700802545543.

[12] K. M. C. Tjørve, E. Tjørve, The use of gompertz models in growth analyses, and new gompertz-model approach: An addition to the unified-richards family, PLoS ONE 12 (6) (2017) 1–17. doi:10.1371/journal.pone.0178691.

[13] D. Yang, P. Gao, C. Tian, Y. Sheng, Gompertz tracking of the growth trajectories of the human-liver-cancer xenograft-tumors in nude mice, Computer Methods and Programs in Biomedicine 191 (2020) 1055412. doi:10.1016/j.cmpb.2020.105412.

[14] E. Tjørve, K. M. C. Tjørve, A unified approach to the richards-model family for use in growth analyses: Why we need only two model forms, Journal of Theoretical Biology 267 (3) (2010) 417–25. doi:10.1016/j.jtbi.2010.09.008.

[15] G. Albano, V. Giorno, P. Román-Román, F. Torres-Ruiz, On the effect of a therapy able to modify both the growth rates in a gompertz stochastic model, Mathematical Biosciences 245 (1) (2013) 12–21. doi:10.1016/j.mbs.2013.01.001.

[16] V. Giorno, A. G. Nobile, Restricted gompertz-type diffusion processes with

periodic regulation functions, Mathematics 7 (6) (2019) 555. doi:10.3390/math7060555.

[17] G. Ascione, E. Pirozzi, On the construction of some fractional stochastic gompertz models, Mathematics 8 (1) (2020) 60. doi:10.3390/math8010060.

[18] L. E. B. Cabrales, J. I. Montijano, M. Schonbek, A. R. S. Castañeda, A viscous modified gompertz model for the analysis of the kinetics of tumors under electrochemical therapy, Mathematics and Computers in Simulation 151 (2018) 96–110. doi:10.1016/j.matcom.2018.03.005.

[19] E. Shakeri, G. Latif-Shabgahi, A. E. Abharian, Predictive drug dosage control through a fokker-planck observer, Computational and Applied Mathematics 37 (3) (2018) 3813–3831. doi:10.1007/s40314-017-0542-x.

[20] G. Albano, V. Giorno, A stochastic model in tumor growth, Journal of Theoretical Biology 242 (2) (2006) 329–336. doi:10.1016/j.jtbi.2006.03.001.

[21] G. Albano, V. Giorno, P. Román-Román, F. Torres-Ruiz, Inference on a stochastic two-compartment model in tumor growth, Computational Statistics and Data Analysis 56 (2012) 1723–1736. doi:10.1016/j.csda.2011.10.016.

[22] G. Albano, V. Giorno, P. Román-Román, F. Torres-Ruiz, Inferring the effect of therapy on tumors showing stochastic Gompertzian growth, Journal of Theoretical Biology 276 (1) (2011) 67–77. doi:10.1016/j.jtbi.2011.01.040.

[23] G. Albano, V. Giorno, P. Román-Román, S. Román-Román, F. Torres-Ruiz, Estimating and determining the effect of a therapy on tumor dynamics by a modified gompertz diffusion process, Journal of Theoretical Biology 364 (7) (2015) 206–219. doi:10.1016/j.jtbi.2014.09.014.

[24] P. Román-Román, S. Román-Román, J. J. Serrano-Pérez, F. Torres-Ruiz, Modeling tumor growth in the presence of a therapy with an effect on rate growth and variability by means of a modified gompertz diffusion process, Journal of Theoretical Biology 407 (1) (2016) 1–17. doi:10.1016/j.jtbi.2016.07.023.

[25] L. J. Höök, E. Lindströmb, Efficient computation of the quasi likelihood function for discretely observed diffusion processes, Computational Statistics and Data Analysis 103 (2016) 426–437. doi:10.1016/j.csda.2016.05.014.

[26] K. Ignatieva, E. Platen, Estimating the diffusion coefficient function for a diversified world stock index, Computational Statistics and Data Analysis 56 (2012) 1333–1349. doi:10.1016/j.csda.2011.10.004.

[27] A. Mandal, W. T. Huang, S. K. Bhandari, A. Basu, Goodness-of-fit testing in growth curve models: A general approach based on finite differences, Computational Statistics and Data Analysis 55 (2011) 1086–1098. doi:10.1016/j.csda.2010.09.003.

[28] A. C. Guidoum, K. Boukhetala, Sim.diffproc: Simulation of diffusion processes. r package version 4.6, https://CRAN.R-project.org/package=Sim.DiffProc.

[29] S. M. Iacus, Simulation and inference for stochastic differential equations, Springer, 2008.

[30] P. Román-Román, S. Román-Román, J. J. Serrano-Pérez, F. Torres-Ruiz, Fitting real data by means of non-homogeneous lognormal diffusion processes, Statistics and Its Interface 10 (4) (2017) 585–600. doi:10.4310/SII.2017.v10.n4.a5.

[31] S. Sheather, M. Jones, A reliable data-based bandwidth selection method for kernel density estimation, Journal of the Royal Statistical Society series B 53 (1991) 683–690. doi:10.1111/j.2517-6161.1991.tb01857.x.

[32] R. Ozols, B. Bundy, B. G. et al., Phase iii trial of carboplatin and paclitaxel compared with cisplatin and paclitaxel in patients with optimally resected stage iii ovarian cancer: A gynecologic oncology group study, Journal of Clinical Oncology 21 (17) (2003) 3194–3200. doi:10.1200/JCO.2003.02.153.

Paper: **Inference on an heteroscedastic Gompertz tumor growth model**

Authors: **Albano, Giorno, Roman, Roman, Serrano and Torres**

Declarations of interest: none