



HAL
open science

Tumor Growth Modeling: Parameter estimation with Maximum Likelihood methods

Spyridon Patmanidis, Alexandros C. Charalampidis, Ioannis Kordonis,
Georgios Mitsis, George P. Papavassilopoulos

► **To cite this version:**

Spyridon Patmanidis, Alexandros C. Charalampidis, Ioannis Kordonis, Georgios Mitsis, George P. Papavassilopoulos. Tumor Growth Modeling: Parameter estimation with Maximum Likelihood methods. *Computer Methods and Programs in Biomedicine*, In press, 10.1016/j.cmpb.2018.03.014 . hal-01746012v1

HAL Id: hal-01746012

<https://hal.science/hal-01746012v1>

Submitted on 1 May 2018 (v1), last revised 22 Sep 2018 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Tumor Growth Modeling: Parameter estimation with Maximum Likelihood methods

Spyridon Patmanidis^{a,*}, Alexandros C. Charalampidis^{b,c}, Ioannis Kordonis^c,
Georgios D. Mitsis^d, George P. Papavassilopoulos^a

^a*School of Electrical and Computer Engineering, National Technical University of Athens, Iroon Polytechniou 9, Zografou 15780, Athens, Greece*

^b*Department of Electrical Engineering and Computer Science, Technische Universität Berlin, Control Systems Group, Einsteinufer 17, Berlin D-10587, Germany*

^c*CentraleSupélec, Automatic Control Group - IETR, Avenue de la Boulaie, 35576 Cesson-Sévigné, France*

^d*Department of Bioengineering, McGill University, 817 Sherbrooke Ave W, MacDonald Engineering Building 270, Montréal QC H3A 0C3, Canada*

Abstract

Background & Objective: In this work, we focus on estimating the parameters of the widely used Gompertz tumor growth model, based on measurements of the tumor's volume. Being able to accurately describe the dynamics of tumor growth on an individual basis is very important both for growth prediction and designing personalized, optimal therapy schemes (e.g. when using model predictive control).

Methods: Our analysis aims to compute both the growth rate and the carrying capacity of the Gompertz function, along with the characteristics of the additive Gaussian process and measurement noise of the system. Three methods based on Maximum Likelihood estimation are proposed. The first utilizes an assumption regarding the measurement noise that simplifies the problem, the second combines the Extended Kalman Filter and Maximum Likelihood estimation, and the third is a nonstandard exact form of Maximum Likelihood estimation,

*Corresponding author: Spyridon Patmanidis.

Address: ECE Building, Office: 21.42, Iroon Polytechniou 9, Zografou 15780, Athens, Greece
Tel: +302107722545, Mob: +306972818983

Email addresses: spatmanid@gmail.com (Spyridon Patmanidis),
alexandros.charalampidis@centralesupelec.fr (Alexandros C. Charalampidis),
jkordonis1920@yahoo.com (Ioannis Kordonis), georgios.mitsis@mcgill.ca (Georgios D. Mitsis), yorgos@netmode.ece.ntua.gr (George P. Papavassilopoulos)

where numerical integration is used to approximate the likelihood of the measurements, along with a novel way to reduce the required computations.

Results: Synthetic data were used in order to perform extensive simulations aiming to compare the methods' effectiveness, with respect to the accuracy of the estimation. The proposed methods are able to estimate the growth dynamics, even when the noise characteristics are not estimated accurately. Another very important finding is that the methods perform best in the case that corresponds to the problem needed to be solved when dealing with experimental data.

Conclusion: Using nonstandard, problem specific techniques can improve the estimation accuracy and best exploit the available data.

Keywords: Tumor Growth Modeling, Nonlinear Systems, Parameter Estimation, Maximum Likelihood, Extended Kalman Filter.

1. Introduction

Cancer is a disease that can affect most of the tissues in the human body. It is defined by uncontrolled cell growth and the potential to invade surrounding tissues. As an additional feature, many cancers, but not all, can also experience
5 migration of tumor cells from the primary site to a distant one where they settle, a phenomenon called metastasis. Cancer also includes benign tumors that show uncontrolled cell growth but no invasion, and in most cases do not threaten the patient's life [1]. According to [2], cancer is a major cause of morbidity and mortality, affecting populations in all countries and regions. The predicted global
10 cancer burden is expected to exceed 20 million new cases annually by 2025, compared with the estimated 14.1 million cases worldwide in 2012. Treatment depends on the type of cancer and commonly consists of one or a combination of therapies, such as surgery, chemotherapy, radiotherapy, immunotherapy and hormone therapy.

15 Mathematical modeling combined with optimal control techniques have been

proposed to improve the outcome of chemotherapy or radiation therapy [3–6] predict the success or failure of immunotherapy [7], and illustrate the potential synergism between different treatment methods [8]. Recent findings also suggest that mathematical modeling can provide useful clues about the impact of surgery
20 on metastasis and may help to predict the risk of cancer spread [9]. An extensive literature review on the topic of how mathematical modeling can be applied to deliver better drug administration regimes is presented in [10].

Tumor growth modeling is a topic that has been studied extensively. The authors of [11] present a concise history of the study of solid tumor growth
25 by reviewing some of the important mathematical contributions since the early decades of the twentieth century. More recent studies about mathematical models that are used to describe tumor growth can be found in [12–15]. In the present study, the Gompertz model is utilized to describe tumor growth. The Gompertzian equation originates from the actuarial model developed by Benjamin Gompertz in 1825 [16], and has been also applied to the study of growth
30 in biological and economic contexts [17]. It has been shown that the Gompertz model can describe the growth of several tumor types [18]. This is also indicated by several recent studies where the Gompertz model is compared with other classical mathematical models (e.g. exponential, Von Bertalanffy, power law, etc) [15, 19, 20]. Therefore, it is a frequent choice for modeling tumor
35 growth in recent experimental studies [3–5], [21–23].

According to the Gompertz model, the growth of a solid tumor is characterized by two main features, the growth rate and the carrying capacity. The growth rate refers to the time needed for the tumor cells to produce new cells,
40 while the carrying capacity to the fact that a tumor cannot grow indefinitely. When fitting experimental tumor data to the Gompertz equation, several studies [22–24] have used the findings of Brunton and Wheldon [25–27], that the carrying capacity of a tumor is a species-specific parameter. However, the theoretical assumption that any given tumor will approach the same asymptotic
45 value has not been proven experimentally.

Tumor growth modeling is inherently difficult as cancer is a complex bio-

logical phenomenon. The fact that its presentation, development and outcome varies from one patient to the other necessitates the need of tumor growth modeling to be done on an individualized basis, in order to better understand the underlying biology and design better therapies. Towards this direction, non-standard, problem specific techniques can provide very accurate models and improve the modeling procedure. An accurate patient-specific growth model combined with mathematical models that simulate the pharmacokinetic and pharmacodynamic processes of drugs and optimal control techniques, can help to predict and optimize the patient's response to various dose schedules and treatment combinations. Moreover, since clinical trials are limited due to their high cost, combining mathematical modeling with numerical simulations is a low cost and safe (since there is no immediate threat for the patient) way to evaluate different strategies.

In this study, it is assumed that both the growth rate and the carrying capacity are unknown (the proposed methods, could also be used if the carrying capacity is considered a species-specific parameter). Using the Gompertz equation, a stochastic state-space model with additive Gaussian process and measurement noise is created. Based on this stochastic model, we try to estimate the growth rate, the carrying capacity and the noise characteristics of tumors. Synthetic data are used because they provide the major advantage that the methods' ability to estimate the unknown parameters can be evaluated objectively. The methods we use to estimate the unknown parameters are based on Maximum Likelihood estimation [28–30]. According to Kay [28], this technique is overwhelmingly the most popular approach to obtain practical estimators. Additionally, the successful application of Maximum Likelihood estimation on problems within a biological or economic context where the estimation of the unknown parameters of stochastic Gompertz-type systems is required [31–33], suggests that it can also be successfully applied for the study of tumor growth.

The main focus of this work is to examine the cases where the characteristics of one or both the process and measurement noise are unknown. A prior work for the case where the characteristics of both process and measurement noise

are known has already been published (see [34]). For the sake of completeness, we also include this case here. Three methods are described in this manuscript.

80 The first will be referred to as Naive Maximum Likelihood (NaML). In this method, a simplification considering the measurement noise has been made in order to obtain the system's states and avoid solving time consuming recursive equations. The second method uses the Extended Kalman Filter (EKF) [35–37], to compute estimates of the system's states and Maximum Likelihood estimation

85 to compute the parameters that maximize the joint density of the observations, and will be referred to as EKF–ML. The third one, which will be referred to as Numerical Maximum Likelihood (NuML), is the method that overall yielded the best results. This method uses numerical integration to compute the integrals that compose the joint density function. We also propose an effective way to

90 reduce the interval of integration and thus the method's execution time, without any significant approximation error. Overall, we were able to achieve good model prediction and accurate parameter estimation.

In the following sections, we present the model used to describe the tumor growth (Section 2), analyze the proposed methods (Section 3.1) and the test

95 cases (Section 3.2), describe the simulations and present their results (Section 4), as well as explain our findings (Section 5) and discuss future work (Section 6).

2. Tumor Growth Model

The Gompertz model has been widely used because of its simplicity and its ability to describe experimental data reasonably well. The discrete time state-space representation of the model is given by:

$$x_{k+1} = f(x_k, \theta_1, \theta_2) = \theta_2 \exp \left(\log \left(\frac{x_k}{\theta_2} \right) \exp \left(- \frac{1}{\theta_1} T \right) \right), \quad (1)$$

where x_k (mm³) is the tumor size, θ_1 (days) is a constant related to the proliferative ability of the cells, θ_2 (mm³) is the carrying capacity ($\lim_{k \rightarrow \infty} x_k = \theta_2$),

100 T (days) is the time interval between k and $k + 1$, and $k \in \mathbb{N}$ [38]. For simplicity reasons, throughout this work, a constant sampling period T is assumed.

Nonetheless, the results of the following sections can be applied to the case of nonconstant sampling intervals.

Assuming random additive process and measurement noise the model can be formulated as follows:

$$x_{k+1} = f(x_k, \theta_1, \theta_2) + w_k, \quad (2a)$$

$$y_k = x_k + v_k, \quad (2b)$$

where y_k is a measurement taken at time step k . The random variables w_k and v_k , $k \in \mathbb{N}$ are mutually independent and normally distributed with zero mean and unknown standard deviation: $w_k \sim \mathcal{N}(0, \theta_3 x_k^{\theta_4})$ and $v_k \sim \mathcal{N}(0, \theta_5 x_k^{\theta_6})$. The notation $X \sim \mathcal{N}(\mu, \sigma)$ is used in this manuscript to denote a normally distributed random variable X with mean μ and standard deviation σ . Measurements are available from time $k=1$ onwards and known x_0, y_0 are assumed. Because of the random components w_k and v_k , the model of Eq. (2) also corresponds to the description:

$$x_{k+1} \sim p_{\theta_a}(x_{k+1}|x_k), \quad (3a)$$

$$y_k \sim p_{\theta_b}(y_k|x_k), \quad (3b)$$

105 where $p_{\theta_a}(x_{k+1}|x_k)$ is the probability density function describing the dynamics for given values of x_k and $p_{\theta_b}(y_k|x_k)$ is the probability density function describing the measurements [30],[35]. In Eq. (3a), $\theta_a = [\theta_1; \theta_2; \theta_3; \theta_4]^T$, where $\theta_a \in \Theta_a$ with $\Theta_a \subseteq \mathbb{R}_{>0}^4$ denoting a compact set of permissible values of the unknown vector θ_a and in Eq. (3b), $\theta_b = [\theta_5; \theta_6]^T$, where $\theta_b \in \Theta_b$ with $\Theta_b \subseteq \mathbb{R}_{>0}^2$
 110 denoting a compact set of permissible values of the unknown vector θ_b . We also define the vector $\theta = \theta_a \times \theta_b$, where $\theta \in \Theta$ with $\Theta \subseteq \mathbb{R}_{>0}^6$

3. Methods

The problem addressed in this section is to obtain an estimate $\hat{\theta}$ based on N measurements $\mathbf{Y}_N = [y_1, \dots, y_n]$. However, as the value of the state variable x_k
 115 is not directly measurable, Maximum Likelihood estimation cannot be applied

directly to Eq. (2b) and Eq. (2b). To overcome this difficulty, in Section (3.1), we propose three methods. The first is a naive approach where, for the sake of simplicity, it is assumed that there is no measurement noise and thus the states x_k equal the measurements y_k . In the second approach, an Extended Kalman
120 Filter is used to estimate the unknown states x_k and then Maximum Likelihood to estimate the system's parameters. Lastly, the third approach, decomposes the likelihood of the observations and approximates it by using one-dimensional recursive numerical integration.

3.1. Parameter Estimation Techniques

125 3.1.1. Naive Maximum Likelihood

In this method, because of the assumption that there is no measurement noise, Maximum Likelihood estimation can be applied directly, as if the measurements were taken from a system of the form:

$$x_{k+1} = f(x_k, \theta_1, \theta_2) + d_k, \quad (4a)$$

$$y_k = x_k, \quad (4b)$$

where $d_k \sim \mathcal{N}(0, \sigma_k)$ and $\sigma_k = \theta_3 x_k^{\theta_4} + \theta_5 x_k^{\theta_6}$. $\hat{\theta}$ is the estimate of the vector θ that maximizes the joint density (likelihood) $p_{\theta}(\mathbf{Y}_N)$ of the observations:

$$\hat{\theta} = \arg \max_{\theta \in \Theta} p_{\theta}(\mathbf{Y}_N). \quad (5)$$

Using Bayes' rule, the likelihood $p_{\theta}(\mathbf{Y}_N)$ can be decomposed according to

$$p_{\theta}(\mathbf{Y}_N) = \prod_{k=1}^N p_{\theta}(y_k | \mathbf{Y}_{k-1}) = \prod_{k=1}^N p_{\theta}(x_k | x_{k-1}), \quad (6)$$

where x_0 is considered known and $p_{\theta}(x_k | x_{k-1})$ is coming from Eq. (4a).

3.1.2. Extended Kalman Filter – Maximum Likelihood

In this method, the EKF is used to provide an approximation \hat{x}_k of the state x_k . In order for the filter to approximate the non-linearities of the system dynamics, a linearized version of the nonlinear system model around the last state estimate is created. Once the state estimates $\hat{\mathbf{X}}_N = [\hat{x}_1, \dots, \hat{x}_n]$ are

available, the maximum likelihood approach can be used to find an estimate $\hat{\boldsymbol{\theta}}_a$ of the vector $\boldsymbol{\theta}_a$, that maximizes the joint density $p_{\boldsymbol{\theta}_a}(\hat{\mathbf{X}}_N)$ of the state estimates:

$$\hat{\boldsymbol{\theta}}_a = \arg \max_{\boldsymbol{\theta}_a \in \Theta_a} p_{\boldsymbol{\theta}_a}(\hat{\mathbf{X}}_N), \quad (7)$$

and an estimate $\hat{\boldsymbol{\theta}}_b$ of the vector $\boldsymbol{\theta}_b$, that maximizes the joint density $p_{\boldsymbol{\theta}_b}(\mathbf{Y}_N)$ of the measurements:

$$\hat{\boldsymbol{\theta}}_b = \arg \max_{\boldsymbol{\theta}_b \in \Theta_b} p_{\boldsymbol{\theta}_b}(\mathbf{Y}_N). \quad (8)$$

If Bayes' rule is applied, the likelihoods $p_{\boldsymbol{\theta}_a}(\hat{\mathbf{X}}_N)$ and $p_{\boldsymbol{\theta}_b}(\mathbf{Y}_N)$ can be decomposed according to

$$p_{\boldsymbol{\theta}_a}(\hat{\mathbf{X}}_N) = \prod_{k=1}^N p_{\boldsymbol{\theta}_a}(\hat{x}_k | \hat{x}_{k-1}) \quad (9)$$

and

$$p_{\boldsymbol{\theta}_b}(\mathbf{Y}_N) = \prod_{k=2}^N p_{\boldsymbol{\theta}_b}(y_k | \hat{x}_k), \quad (10)$$

where $p_{\boldsymbol{\theta}_a}(\hat{x}_k | \hat{x}_{k-1})$ comes from Eq. (2a), $p_{\boldsymbol{\theta}_b}(y_k | \hat{x}_k)$ from Eq. (2b) and \hat{x}_0 is considered known.

130 The EKF used in this work is an implementation of the classical EKF [36].
 Moreover, it has been observed that when this procedure is used repeatedly, $\hat{\boldsymbol{\theta}}$
 will converge to a better estimate than the initial one. The reason behind this
 behavior is that the vector $\hat{\boldsymbol{\theta}}$ is considered known for the EKF and the state
 estimation is based on this $\hat{\boldsymbol{\theta}}$. However, when the state estimates are used to
 135 maximize Eq. (9) and (10) the updated $\hat{\boldsymbol{\theta}}$ is a better estimate of $\boldsymbol{\theta}$ and when
 the EKF is used again with this new $\hat{\boldsymbol{\theta}}$, the new states \hat{x}_k will be improved
 estimates of the real states x_k , and so on. In the first iteration, the estimate
 $\hat{\boldsymbol{\theta}}$ that results from NaML is used for the EKF, and in the following iterations
 the latest estimate $\hat{\boldsymbol{\theta}}$ of Eq. (7) and (8) is used. This procedure stops when two
 140 consecutive estimates do not diverge over a specific threshold or after a preset
 number of iterations has been reached.

3.1.3. Numerical Maximum Likelihood

This method is applied to the model described in Eq. (2) and is used to find an estimate $\hat{\boldsymbol{\theta}}$ of the vector $\boldsymbol{\theta}$ that maximizes the joint density $p_{\boldsymbol{\theta}}(\mathbf{Y}_N)$ of the observation:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} p_{\boldsymbol{\theta}}(\mathbf{Y}_N). \quad (11)$$

Bayes' rule can be used in order to decompose the joint density according to

$$p_{\boldsymbol{\theta}}(\mathbf{Y}_N) = p_{\boldsymbol{\theta}}(y_1) \prod_{k=2}^N p_{\boldsymbol{\theta}}(y_k | \mathbf{Y}_{k-1}), \quad (12)$$

where

$$p_{\boldsymbol{\theta}}(y_{k+1} | \mathbf{Y}_k) = \int p_{\boldsymbol{\theta}}(y_{k+1} | x_{k+1}) p_{\boldsymbol{\theta}}(x_{k+1} | \mathbf{Y}_k) dx_{k+1}, \quad (13)$$

$$p_{\boldsymbol{\theta}}(x_{k+1} | \mathbf{Y}_k) = \int p_{\boldsymbol{\theta}}(x_{k+1} | x_k) p_{\boldsymbol{\theta}}(x_k | \mathbf{Y}_k) dx_k, \quad (14)$$

and

$$p_{\boldsymbol{\theta}}(x_{k+1} | \mathbf{Y}_{k+1}) = \frac{p_{\boldsymbol{\theta}}(y_{k+1} | x_{k+1}) p_{\boldsymbol{\theta}}(x_{k+1} | \mathbf{Y}_k)}{p_{\boldsymbol{\theta}}(y_{k+1} | \mathbf{Y}_k)}. \quad (15)$$

The distributions in Eq. (13) to (15) can be computed iteratively. Numerical integration can be used to compute the integrals of Eq. (13) and (14). The interval of integration is $[0, X_{max}]$, where X_{max} is the maximum volume a tumor can attain, and the definite integrals that come of, can be computed by using the trapezoidal rule:

$$\int_a^b f(x) dx \approx (b-a) \left[\frac{f(a) + f(b)}{2} \right]. \quad (16)$$

In Eq. (16), the smaller the interval $[a, b]$ is, the smaller the approximation error will be. The integral defined in $[0, X_{max}]$ can be divided into subintegrals, so
145 that the total integral comes as the summation of the individual subintegrals. However, due to the size of $[0, X_{max}]$ and the number of subintegrals needed for an accurate approximation, such an approach is completely ineffective, because making a huge number of calculations at every iteration is prohibitive. An effective way to limit the interval of integration is of great importance, because
150 it reduces the number of subintegrals needed for the approximation and thus the execution time.

Using the Gaussian property of the noise, the interval of integration can be reduced in order to make the numerical integration by using the trapezoidal rule applicable. At every iteration step, if x_k is considered a normally distributed random variable with mean $\mu = y_k$ and standard deviation $\sigma = \sigma_0 y_k^{e_0}$ ($x_k \sim \mathcal{N}(\mu, \sigma)$), then $[\mu - 5\sigma, \mu + 5\sigma] \cap [0, X_{max}]$ can be used as the interval of integration, which is significantly smaller than $[0, X_{max}]$. In this manuscript, the maximum obtainable values of θ_5 and θ_6 are used as σ_0 and e_0 respectively. The integral defined in the new interval can be divided in subintegrals, that can be approximated by applying the trapezoidal rule. As a result, the distributions of Eq. (13) to (15) can be approximated very accurately.

3.2. Test Cases

In this work, we present four test cases. In all cases, the system parameters are unknown. The test cases differ in terms of the available information about the process and measurement noise characteristics. Case 1 has been studied in [34] and is included in order to be compared with the three new cases.

3.2.1. Case 1 - Known Noise Characteristics

In this case, the unknown parameters are the growth rate and the carrying capacity (θ_1, θ_2) , while the characteristics of the process (θ_3, θ_4) and measurement noise (θ_5, θ_6) are considered known. This is the simplest case, because only two parameters need to be estimated. Parameters $\theta_3, \theta_4, \theta_5$ and θ_6 are set to their real values and the argument that maximizes the likelihood $p_{\theta}(\mathbf{Y}_n)$ is $\hat{\theta} = [\hat{\theta}_1; \hat{\theta}_2]^T$.

3.2.2. Case 2 - Known Process Noise Characteristics

In this case, the unknown parameters are $\theta_1, \theta_2, \theta_5$ and θ_6 , while θ_3 and θ_4 are considered known. Parameters θ_3 and θ_4 are set to their real values and the argument that maximizes the likelihood $p_{\theta}(\mathbf{Y}_n)$ is $\hat{\theta} = [\hat{\theta}_1; \hat{\theta}_2; \hat{\theta}_5; \hat{\theta}_6]^T$.

3.2.3. Case 3 - Known Measurement Noise Characteristics

In this case, the unknown parameters are $\theta_1, \theta_2, \theta_3$ and θ_4 , while θ_5 and θ_6 are considered known. Parameters θ_5 and θ_6 are set to their real values and the argument that maximizes the likelihood $p_{\theta}(\mathbf{Y}_n)$ is $\hat{\theta} = [\hat{\theta}_1; \hat{\theta}_2; \hat{\theta}_3; \hat{\theta}_4]^T$.

3.2.4. Case 4 - Unknown Noise Characteristics

In this case, all parameters ($\theta_1, \theta_2, \theta_3, \theta_4, \theta_5$ and θ_6) are unknown and the techniques are used as described previously in this section.

3.3. Implementation

The computational methods presented in this section, were implemented in MATLAB (ver. R2015a, The MathWorks Inc., Natick, IL) by writing the suitable functions. In order to maximize the likelihoods of Eq. (6), (9), (10) and (12) the function `fmincon` [39] was used to find the argument that minimizes the negative logarithm of the likelihood $p_{\theta}(\text{Data})$:

$$\hat{\theta} = \arg \min_{\theta \in \Theta} \left(-\log(p_{\theta}(\text{Data})) \right), \quad (17)$$

where *Data* stands in for \mathbf{Y}_N or $\hat{\mathbf{X}}_N$ depending on the applied method and the parameters that need to be estimated. Maximizing the log-likelihood is equivalent since the logarithm is a monotonic function. It has also been proved numerically using the definition of convexity (e.g. [40]), that the likelihood functions of this manuscript are generally not convex. In order to reduce the possibility of getting stuck to a local minimum, we ran `fmincon` for many different random initial conditions. The simulations were performed using an Intel Core i7-6700K @ 4.00GHz and 16GB of DDR4 @ 3200 MHz.

4. Simulations & Results

In this section, we present the results from the simulations in MATLAB. The growth of 100 cancer tumors was simulated using Eq. (2a) and (2b). Each tumor's growth rate θ_1^j and its carrying capacity θ_2^j , as well as the process and measurement noise characteristics $\theta_3^j, \theta_4^j, \theta_5^j, \theta_6^j$ ($j = 1, \dots, 100$ is the identification

number of the test subject) were chosen randomly from uniform distributions.
 200 The minimum and maximum values for the parameters (see Table 1) were chosen
 so that the obtained dataset appears as realistic tumor measurements. The
 sampling time between two consecutive measurements is $T = 2$ (days) and the
 number of available measurements is $N = 30$.

Table 1

Minimum and maximum values for each parameter

Parameters	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
Minimum	6	320	1	0.3	2	0.2
Maximum	16	800	4	0.5	6	0.5

In order to evaluate the performance of the methods described in Section
 3.1 and applied to the cases of Section 3.2, the Mean Absolute Percentage Error
 (MAPE) and the standard deviation of the Absolute Percentage Errors (APE)
 are used. The MAPE is given by:

$$\text{MAPE} = \frac{1}{j} \sum_{i=1}^j \text{APE}_i(\%), \quad (18)$$

and the standard deviation by:

$$\text{STD} = \left(\frac{1}{j-1} \sum_{i=1}^j |\text{APE}_i - \text{MAPE}|^2 \right)^{1/2}. \quad (19)$$

In Eq. (18) and (19)

$$\text{APE}_i = \left| \frac{A_i - E_i}{A_i} \right|, \quad (20)$$

where A_i is the actual value of the parameter, E_i is the estimated value and
 205 $j(= 100)$ is the number of the tumors.

As mentioned in Section 3.3, regarding the minimization of Eq. (17), `fmincon`
 has been executed for many different random initial conditions. For all the ran-
 dom initial condition, the difference between the parameter estimates is very
 small, and especially for parameters θ_1 and θ_2 is practically negligible.

210 Fig. 1 presents MAPE and the standard deviation of the APE for the growth
 rate and the carrying capacity at each iteration step (each time a new measure-

ment is added to the system) for each test case of Section 3.2 that were obtained with the NuML. A comparison between the different methods is presented for case 4 only, however similar figures are obtained and the same conclusions hold
215 when comparing the methods to the other test cases. In Fig. 2a and Fig. 2b, the MAPEs of the growth rate and carrying capacity respectively are shown, while Fig. 2c shows the MAPE of the noise parameters. Fig. 3 shows the STD for the previously mentioned parameters for all the estimation methods used. The three vertical lines show the mean time needed to reach the 50%
220 (about 7 measurements), 75% (about 11 measurements) and 100% (about 21 measurements) of the carrying capacity respectively.

Fig. 4 shows the tumor growth progression for two randomly chosen test subjects, and the ability of Gompertz curves to fit the data. These curves are created at different time-steps, meaning that the number of available measurements differs. The unknown parameters were estimated with the NuML, which
225 is the method that yielded the best results. However, both NaML and EKF-ML yield similar results. The red curve is created by the parameters estimated from all the measurement. The parameters used for the green curve are reestimated for every new measurement (practically, the one-step ahead prediction
230 of the growth). For the blue curve, the estimate is obtained by the three initial measurements. Likewise, for the magenta, the cyan and the black curve, the measurements until the 50%, 75% and 100% of the carrying capacity respectively are used in order to obtain an estimate for the unknown parameters and predict future growth.

235 5. Discussion

Despite the large number of studies in tumor growth, the widely used theoretical assumption that the carrying capacity is a species-specific parameter has not been proved experimentally. On the other hand, even though there is strong evidence that tumor growth is subject-specific since trying to estimate
240 the carrying capacity may lead to unrealistic results, it is preferred to keep a

fixed value for this parameter [22]. In this manuscript, our simulation results show that the methods proposed in Section 3.1 can estimate accurately both the growth rate and the carrying capacity. Realistic estimates for the carrying capacity are obtained even when the number of available measurements is very
245 small.

Four different cases have been studied depending on the available information considering the noise. The metrics we used to evaluate the methods' performance regarding the estimation of the growth dynamics on these cases are presented in Fig. 1. The most important test case is the fourth, where both
250 the process and measurement noise characteristics are unknown. This case corresponds to the problem needed to be solved when dealing with experimental data. Even when the noise type is known, it is still difficult to know the characteristics of the noise. The other three cases were studied in order to test if knowing the exact values for the noise characteristics could lead in a more
255 accurate estimation of the growth dynamics. However, for all three methods, the results show that most times, knowing the exact values of the noise characteristics does not considerably improve the estimation of the growth rate and carrying capacity. On the contrary, considering the noise characteristics unknown seems to give flexibility in finding estimates for the noise characteristics
260 that describe the noise realization better than the real values, which leads to more accurate estimates for the growth parameters. Eq. (17) seems to be more sensitive to parameters θ_1 and θ_2 and can be further minimized by choosing parameter values for θ_3 , θ_4 , θ_5 and θ_6 , that differ from the real parameters. As a result, the fact that the noise characteristics cannot be estimated accurately
265 does not affect the methods' ability to provide accurate estimates for the growth rate and the carrying capacity.

In this work, three methods have been used to estimate the unknown parameters of the model described by Eq. (2a) and (2b). The simulation results are shown in Fig. 2. Overall, NuML performs better than the other two methods
270 regarding the estimation of growth rate and carrying capacity. As regards the noise characteristics, there was no method that could estimate all the noise pa-

rameters consistently better than the others. In Fig. 2c, we can see that NuML performs better in estimating parameters θ_3 and θ_6 , while EKF-ML is better at estimating θ_4 and θ_5 .

275 Another important finding is that as new measurements become available, the estimates of growth rate and carrying capacity converge to a value that is better than the previous values, but this does not happen for the estimates of the noise characteristics. Of course, after the carrying capacity has been reached, the new measurements do not provide much information about the growth rate, 280 so the growth rate estimates do not improve significantly after that point on. In the same manner, the initial measurements do not provide a lot of information about the carrying capacity. As we can see, the estimation of carrying capacity improves a lot after the 75% of the tumor's maximum size has been reached, and when the tumor has fully grown it can be estimated very accurately. However, 285 even though the initial estimates of the carrying capacity are not very accurate, they still provide realistic values for this parameter.

NaML is not as accurate as NuML and EKF-ML in estimating parameters θ_1 and θ_2 . This is due to the simplification concerning the way noise is added to the system. However, this method is very useful because the estimates are rea- 290 sonably good, and most importantly, the execution time is extremely low (less than 0.15 secs when 30 measurements are available). The NaML estimates can be used as initial conditions for the other methods, as it has been observed that having a good initial guess instead of random one reduces the execution time of NuML and the number of iterations needed for the EKF-ML to converge. Es- 295 pecially for NuML, there are cases where the execution time is reduced by half, when the NaML estimates are used. The EKF-ML method seems to perform better than NaML in estimating the growth rate and carrying capacity. The execution time of this methods is low (about 1.75 secs when 30 measurements are available), but not as low as the execution time of NaML. The NuML method 300 gives the best estimates for the parameters θ_1 and θ_2 compared to NaML and EKF-ML. Even though all three methods can estimate the growth of a tumor very well, the estimates from NuML create curves that fit better to the mea-

sured data. Moreover, the one-step ahead prediction performance has promising potential. However, the execution time of NuML is very high (about 125 secs when 30 measurements are available) compared to the ones from NaML and EKF-ML due to the recursion and the computationally demanding approximations of the integrals, but yet far away from being prohibitive. Fig. 5 shows the mean execution time of EKF-ML and NuML at every iteration step. In our experiments, NuML improved the estimation accuracy of θ_1 and θ_2 by 0.3 to 2.3% and 0.3 to 1.75% respectively compared to EKF-ML. Even though the previous numbers imply that the improvement is small, waiting for a couple of minutes in order to obtain a better estimate should not discourage the use of NuML.

Cancer is a complex biological phenomenon that is affected by various factors and occurs on many levels [41]. The lack of appropriate experimental data makes infeasible to fully validate dynamic cancer models, which would be the most principled approach for capturing tumor growth [22]. Any attempt to model cancer progression has to rely on incomplete information. Due to the fact that no such thing as a universal model for all cancers exists and since all models are bound by limitations, the model selection is usually based on the type of data that are used. In this study, the Gompertz model is utilized to describe the tumor growth. The main criticism of this model is that the relative tumor growth rate becomes arbitrary large for small tumor volumes [15]. Additionally, it does not model any biological mechanism that underly cancer, which means that it does not provide any insight about the biological procedures. It has been shown that it provides best fit for breast and lung cancer [15] and is very popular for large animal tumor datasets, but it has not been shown that it can model accurately all types of cancer. However, it is a simple model that is suitable for macroscopic tumor growth modeling. The fact that it can be used in optimal control algorithms during treatment optimization makes this model very popular. It has been shown to be appropriate for the prediction of the average growth behavior of a tumor and that it fits tumor growth datasets with accuracy [15, 18, 42, 43]. Moreover, the experimental data that we plan

to work with in our future work and on which the synthetic data used here are
335 based are of macroscopic tumor size. In the majority of studies on this type of
data the Gompertz model is used. Thus, we consider that our choice to use this
model is well justified. More information about the advantages and limitations
of some of the most popular tumor growth models and the possible reasons
why the Gompertz model ended up being the most widely applied one, can be
340 found in the work of P. Gerlee [14]. Nevertheless, the model’s selection does not
restrict us from utilizing other models in the future if needed, as the methods
we propose are not limited by the model’s choice.

The evidence that cancer is a subject-specific disease [22] along with obser-
vations that the clinical response of individual patients to therapy still remains
345 uncertain [44] make the need for further research on individualized tumor pre-
diction models more obvious than ever. Moreover, modern computers are able
to run nonstandard, problem specific techniques in reasonable time and provide
accurate personalized tumor prediction models. The methodology we propose
gives promising results towards this direction and can become a very useful tool
350 in cancer treatment research by allowing us to better predict individual tumor
growth and response to treatment.

6. Conclusion

In this work, we developed three methods in order to estimate the unknown
parameters of the Gompertz function and the characteristics of the Gaussian
355 process and measurement noise. Synthetic data were used to test the meth-
ods’ effectiveness and accuracy. It has been found that tumor growth can be
modeled adequately, even when the noise characteristics cannot be estimated
accurately. Simulation results show that, overall, NuML performs better com-
pared to NaML and EKF–ML. NuML can also provide realistic estimates for
360 the growth dynamics, even when a few measurements are available.

During our work with the synthetic data, we observed that applying nonstan-
dard techniques provided improved estimates and better fitting curves compared

to classical estimation techniques (e.g. Least Squares). This work constitutes an important step towards our future research, the application and testing of
365 the methods we propose on experimental cancer data. Even though these techniques are computationally demanding, nowadays the available computational power is more than enough. So, since even the slightest improvement in tumor growth prediction could be really important, we believe that trading speed off for accuracy is completely justified.

370 **Conflict of interest**

The authors of this manuscript declare that they have no financial and personal relationships with other people or organizations that could inappropriately influence their work.

Acknowledgment

375 Spyridon Patmanidis wishes to thank the Special Account for Research of the National Technical University of Athens for supporting his doctoral research through a scholarship of excellence.

The work of Alexandros Charalampidis has been supported by the European Union's Horizon 2020 research and innovation programme under the Marie
380 Skłodowska-Curie grant agreement No 705982.

References

- [1] E. Barillot, L. Calzone, P. Hupe, J. P. Vert, A. Zinovyev, Computational Systems Biology of Cancer, 2nd Edition, CRC Press, Boca Raton, Florida, 2012.
- 385 [2] B. W. Stewart, C. P. Wild, World Cancer Report 2014, International Agency for Research on Cancer, Lyon, France, 2014.

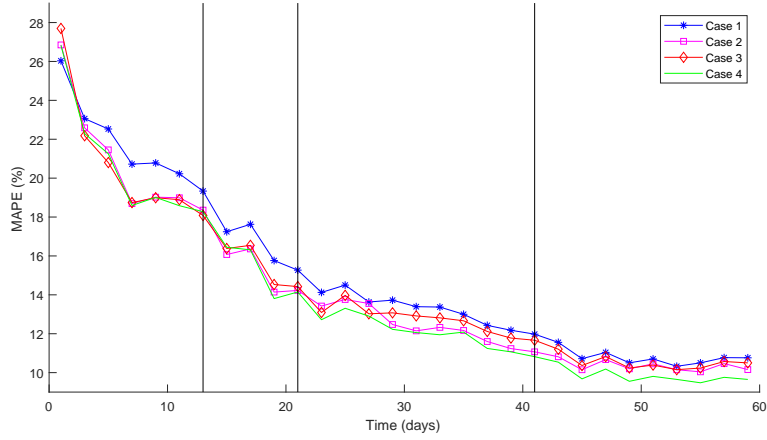
- [3] M. M. Hadjiandreou, G. D. Mitsis, Mathematical modeling of tumor growth, drug-resistance, toxicity, and optimal therapy design, *IEEE Transactions on Biomedical Engineering* 61 (2) (2014) 415–425. doi:10.1109/TBME.2013.2280189.
- 390
- [4] P. Dua, V. Dua, E. N. Pistikopoulos, Optimal delivery of chemotherapeutic agents in cancer, *Computers & Chemical Engineering* 32 (1) (2008) 99–107. doi:https://doi.org/10.1016/j.compchemeng.2007.07.001.
- [5] D. Barbolosi, A. Iliadis, Optimizing drug regimens in cancer chemotherapy: a simulation study using a pk-pd model, *Computers in Biology and Medicine* 31 (3) (2001) 157–172. doi:https://doi.org/10.1016/S0010-4825(00)00032-9.
- 395
- [6] R. C. Rockne, P. Frankel, *Advances in Radiation Oncology*, Springer, Cham, Switzerland, 2017. doi:https://doi.org/10.1007/978-3-319-53235-6_12.
- 400
- [7] C. Babbs, Predicting success or failure of immunotherapy for cancer: insights from a clinically applicable mathematical model, *American Journal of Cancer Research* 2 (2) (2012) 204–213.
- [8] R. Serre, S. Benzekry, L. Padovani, C. Meille, N. André, J. Ciccolini, F. Barlesi, X. Muracciole, D. Barbolosi, Mathematical modeling of cancer immunotherapy and its synergy with radiotherapy, *Cancer Research* 76 (17) (2016) 4931–4940. doi:10.1158/0008-5472.CAN-15-3567.
- 405
- [9] S. Benzekry, A. Tracz, M. Mastro, R. Corbelli, S. Barbolosi, J. M. L. Ebos, Modeling spontaneous metastasis following surgery: An in vivo-in silico approach, *Cancer Research* 76 (3) (2016) 535–547. doi:10.1158/0008-5472.CAN-15-1389.
- 410
- [10] F. Michor, K. Beal, Improving cancer treatment via mathematical modeling: Surmounting the challenges is worth the effort 163 (5) (2015) 1059–1063.

- 415 [11] R. P. Araujo, D. L. S. McElwain, A history of the study of solid tumour growth: The contribution of mathematical modelling, *Bulletin of Mathematical Biology* 66 (5) (2004) 1039–1091. doi:10.1016/j.bulm.2003.11.002.
- [12] H. Enderling, M. Chaplain, Mathematical modeling of tumor growth and
420 treatment, *Current Pharmaceutical Design* 20 (30) (2014) 4934–4940. doi:10.2174/1381612819666131125150434.
- [13] H. Byrne, T. Alarcon, M. Owen, S. Webb, P. Maini, Modelling aspects of cancer dynamics: a review, *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 364 (1843)
425 (2006) 1563–1578. doi:10.1098/rsta.2006.1786.
- [14] P. Gerlee, The model muddle: In search of tumor growth laws, *Cancer Research* 73 (8) (2013) 2407–2411. doi:10.1158/0008-5472.CAN-12-4355.
- [15] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. L. Ebos, L. Hlatky, P. Hahnfeldt, Classical mathematical models for description and prediction
430 of experimental tumor growth, *PLOS Computational Biology* 10 (8). doi:10.1371/journal.pcbi.1003800.
- [16] B. Gompertz, On the nature of the function expressive of the law of human mortality, and on a new mode of determining the value of life contingencies, *Philosophical Transactions of the Royal Society of London* 115 (1825) 513–
435 583.
- [17] C. P. Winsor, The gompertz curve as a growth curve, *Proceedings of the National Academy of Sciences of the United States of America* 18 (1) (1932) 1–8.
- [18] A. K. Laird, Dynamics of tumour growth, *British Journal of Cancer* 18 (3)
440 (1965) 490—502.
- [19] N. Hartung, S. Mollard, D. Barbolosi, A. Benabdallah, G. Chapuisat, G. Henry, S. Giacometti, A. Iliadis, J. Ciccolini, C. Faivre, F. Hubert,

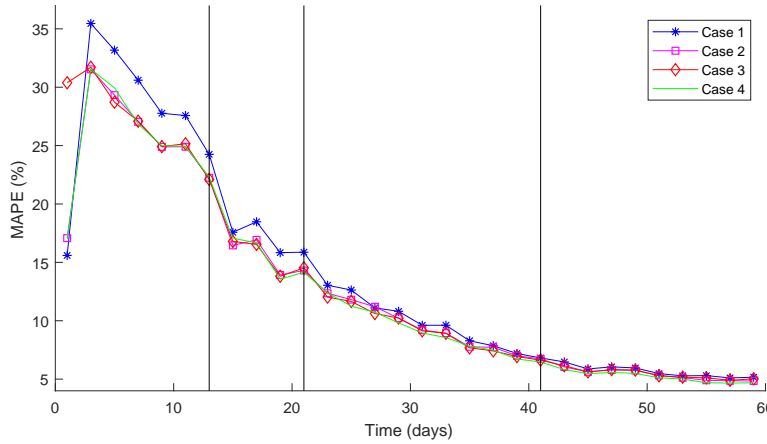
- Mathematical modeling of tumor growth and metastatic spreading: Validation in tumor-bearing mice, *Cancer Research* 74 (22) (2014) 6397–6407. doi:10.1158/0008-5472.CAN-14-0721.
- 445
- [20] E. A. Sarapata, L. G. de Pillis, A comparison and catalog of intrinsic tumor growth models, *Bulletin of Mathematical Biology* 76 (8) (2014) 2010–2024. doi:10.1007/s11538-014-9986-y.
- [21] A. Achilleos, C. Loizides, M. Hadjiandreou, T. Stylianopoulos, G. D. Mitsis, Multiprocess dynamic modeling of tumor evolution with bayesian tumor-specific predictions, *Annals of Biomedical Engineering* 42 (5) (2014) 1095–1111. doi:10.1007/s10439-014-0975-y.
- 450
- [22] C. Loizides, D. Iacovides, M. M. Hadjiandreou, G. Rizki, A. Achilleos, K. Strati, G. D. Mitsis, Model-based tumor growth dynamics and therapy response in a mouse model of de novo carcinogenesis, *PLOS ONE* 10 (12) (2015) 1–18. doi:10.1371/journal.pone.0143840.
- 455
- [23] R. Chignola, R. I. Foroni, Estimating the growth kinetics of experimental tumors from as few as two determinations of tumor size: implications for clinical oncology, *IEEE Transactions on Biomedical Engineering* 52 (5) (2005) 808–815. doi:10.1109/TBME.2005.845219.
- 460
- [24] A. Talkington, R. Durrett, Estimating tumor growth rates in vivo, *Bulletin of Mathematical Biology* 77 (10) (2015) 1934–1954. doi:10.1007/s11538-015-0110-8.
- [25] G. F. Brunton, T. E. Wheldon, Prediction of the complete growth pattern of human multiple myeloma from restricted initial measurements, *Cell Proliferation* 10 (6) (1977) 591–594. doi:10.1111/j.1365-2184.1977.tb00316.x.
- 465
- [26] G. F. Brunton, T. E. Wheldon, Characteristic species dependent growth patterns of mammalian neoplasms, *Cell Proliferation* 11 (2) (1978) 161–175. doi:10.1111/j.1365-2184.1978.tb00884.x.
- 470

- [27] G. F. Brunton, T. E. Wheldon, The gompertz equation and the construction of tumour growth curves, *Cell Proliferation* 13 (4) (1980) 455–460. doi:10.1111/j.1365-2184.1980.tb00486.x.
- [28] S. Kay, *Fundamentals Of Statistical Signal Processing*, Prentice-Hall, Inc.,
475 Upper Saddle River, NJ, USA, 1993.
- [29] L. Ljung, *System Identification: Theory for the User*, Prentice-Hall, Inc.,
Upper Saddle River, NJ, USA, 1986.
- [30] T. B. Schön, A. Wills, B. Ninness, System identification of nonlinear state-space models, *Automatica* 47 (1) (2011) 39–49. doi:https://doi.org/
480 10.1016/j.automatica.2010.10.013.
- [31] M. L. Garg, B. R. Rao, C. K. Redmond, Maximum-likelihood estimation of the parameters of the gompertz survival function, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 19 (2) (1970) 152–159.
- [32] P. Lambert, Modeling of nonlinear growth curve on series of correlated
485 count data measured at unequally spaced times: A full likelihood based approach, *Biometrics* 52 (1) (1996) 50–55.
- [33] R. Gutiérrez, R. Gutiérrez-Sanchez, A. Nafidi, P. Román, F. Torres, Inference in gompertz-type nonhomogeneous stochastic systems by means of discrete sampling, *Cybernetics and Systems* 36 (2) (2005) 203–216.
490 doi:10.1080/01969720590897233.
- [34] S. Patmanidis, A. C. Charalampidis, I. Kordonis, G. D. Mitsis, G. P. Papavassilopoulos, Comparing methods for parameter estimation of the gompertz tumor growth model, *IFAC-PapersOnLine* 50 (1) (2017) 12203–12209. doi:https://doi.org/10.1016/j.ifacol.2017.08.2289.
- [35] A. C. Charalampidis, G. P. Papavassilopoulos, Development and numerical
495 investigation of new non-linear kalman filter variants, *IET Control Theory Applications* 5 (10) (2011) 1155–1166. doi:10.1049/iet-cta.2010.0553.

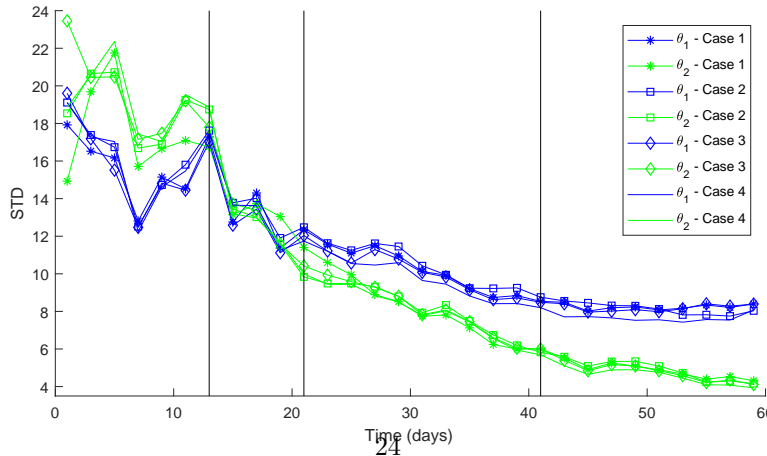
- [36] A. C. Charalampidis, G. P. Papavassilopoulos, Computationally efficient kalman filtering for a class of nonlinear systems, *IEEE Transactions on Automatic Control* 56 (3) (2011) 483–491. doi:10.1109/TAC.2010.2078090.
500
- [37] P. Maybeck, *Stochastic Models, Estimation, and Control*, Mathematics in science and engineering, Academic Press, 1982.
- [38] B. Dennis, J. M. Ponciano, Density-dependent state-space model for population-abundance data with unequal time intervals, *Ecology* 95 (8)
505 (2014) 2069–2076. doi:10.1890/13-1486.1.
- [39] *Matlab Optimization Toolbox User’s Guide*, The MathWorks Inc., 2017.
- [40] D. P. Bertsekas, *Nonlinear Programming*, Athena Scientific, 1999.
- [41] D. Hanahan, R. A. Weinberg, Hallmarks of cancer: The next generation, *Cell* 144 (5) (2011) 646–674. doi:http://dx.doi.org/10.1016/j.cell.
510 2011.02.013.
- [42] T. Wheldon, *Mathematical models in cancer research*, Medical science series, A. Hilger, 1988.
- [43] L. Norton, R. Simon, H. D. Brereton, A. E. Bogden, Predicting the course of gompertzian growth, *Nature* 264 (1976) 542–545. doi:http://dx.doi.
515 org/10.1038/264542a0.
- [44] M. Volm, T. Efferth, Prediction of cancer drug resistance and implications for personalized medicine, *Frontiers in Oncology* 5 (2015) 282. doi:http://doi.org/10.3389/fonc.2015.00282.



(a) MAPE of Growth Rate

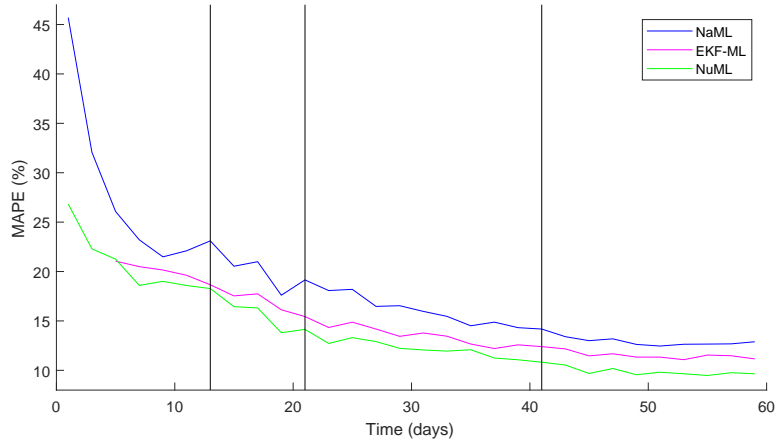


(b) MAPE of Carrying Capacity

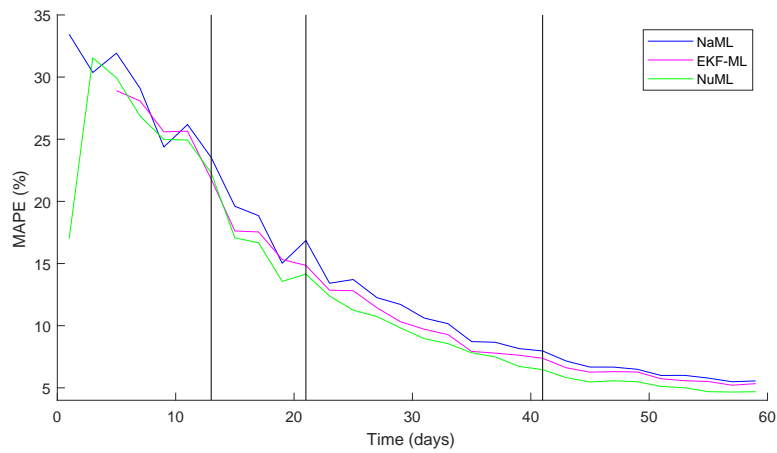


(c) STD of Growth Rate and Carrying Capacity

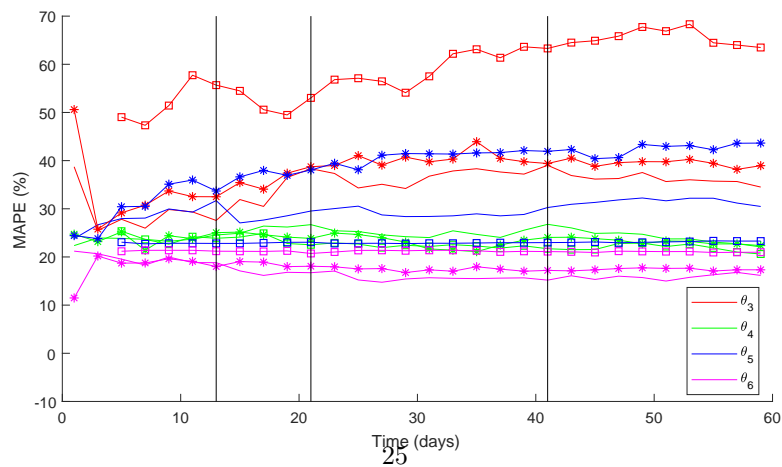
Fig. 1. MAPE and STD of parameters θ_1 and θ_2 for the NuML method. 1a and 1b show the MAPE for θ_1 and θ_2 respectively, while 1c shows the STD of θ_1 and θ_2 .



(a) MAPE of Growth Rate

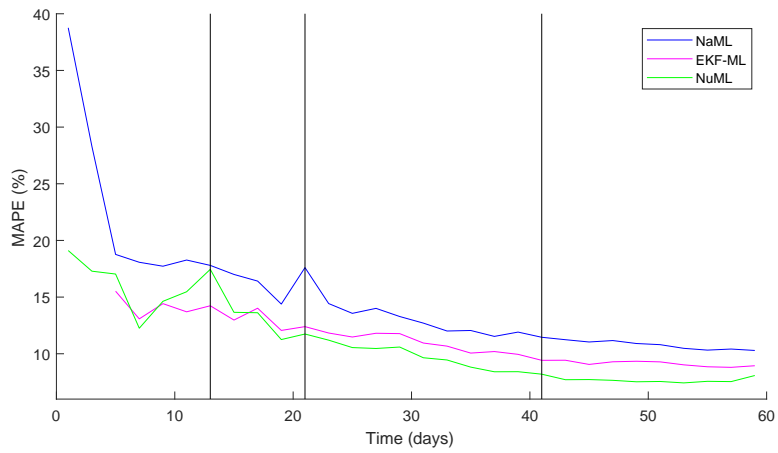


(b) MAPE of Carrying Capacity

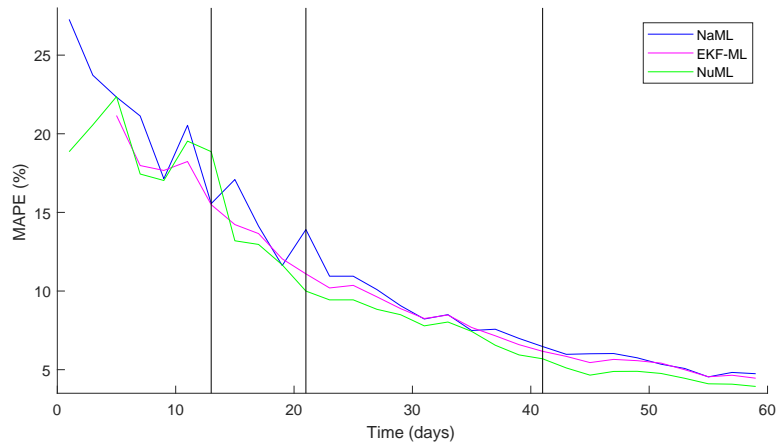


(c) MAPE of Noise Characteristics

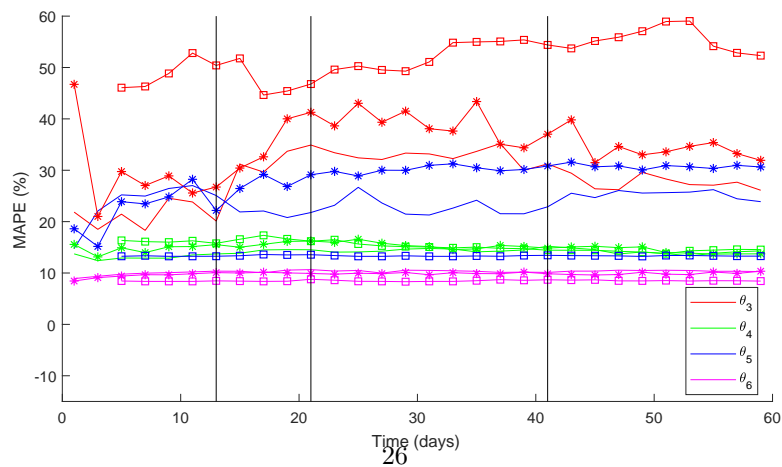
Fig. 2. MAPE of the unknown parameters for test case 4. 2a and 2b show the MAPE of θ_1 and θ_2 respectively. 2c shows the MAPE of θ_3 , θ_4 , θ_5 and θ_6 . In 2c, the line representation refers to the NuML method, the line-asterisk to the NaML method and the line-square to the EKF-ML method.



(a) STD of Growth Rate

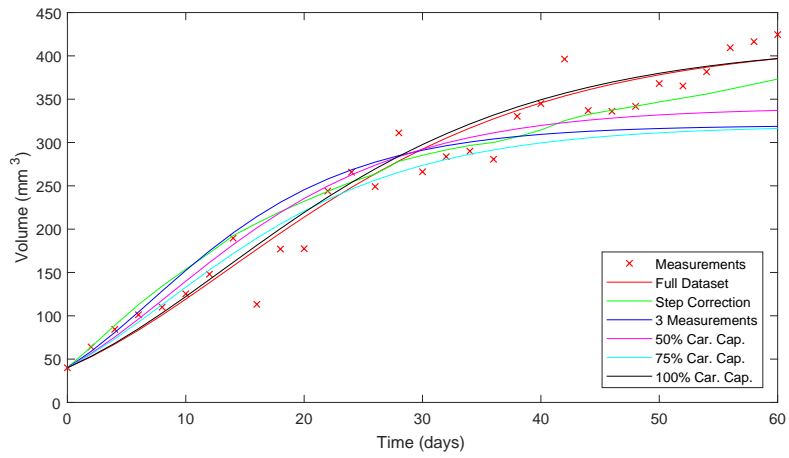


(b) STD of Carrying Capacity

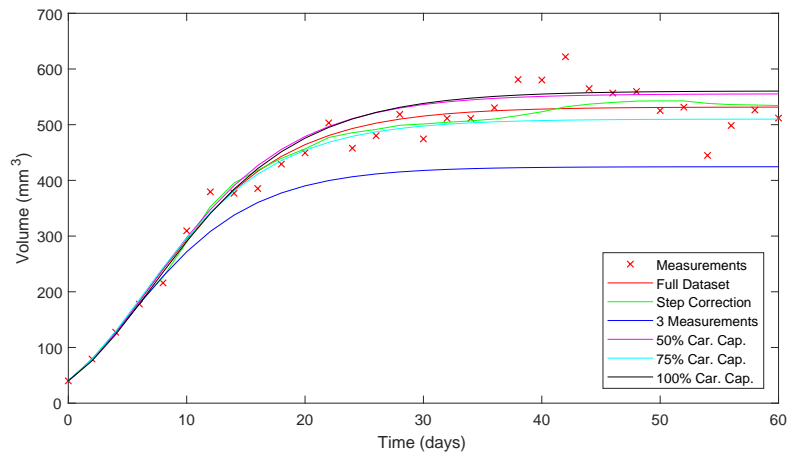


(c) STD of Noise Characteristics

Fig. 3. STD of the unknown parameters for test case 4. 3a and 3b show the STD of θ_1 and θ_2 respectively. 3c shows the STD of θ_3 , θ_4 , θ_5 and θ_6 . In 3c, the line representation refers to the NuML method, the line-asterisk to the NaML method and the line-square to the EKF-ML method.



(a) Tumor 1



(b) Tumor 2

Fig. 4. Growth curves for random test subjects. In each subfigure, the x marks represents y_k while the colored curves describe the tumor growth estimation based on the available measurements. The unknown parameters have been estimated with the NuML.

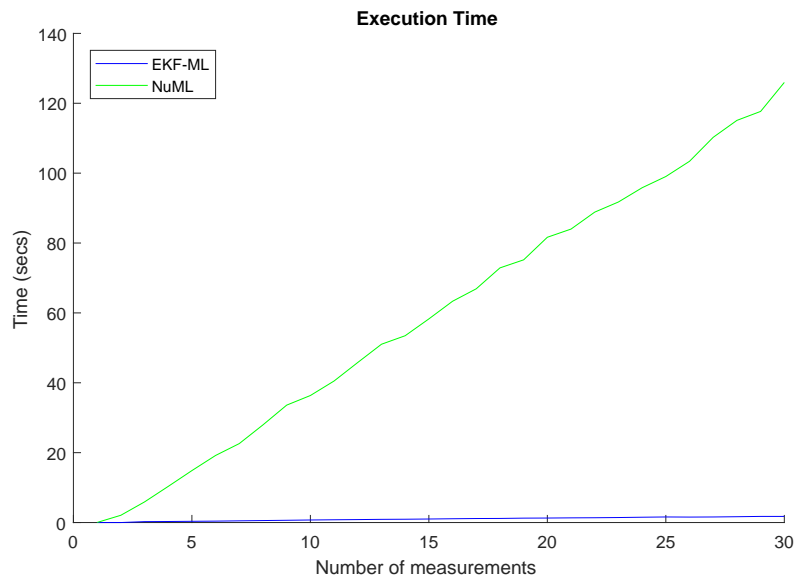


Fig. 5. Execution time for EKF-ML and NuML.