



HAL
open science

A comparison between Nonlinear Least Squares and Maximum Likelihood estimation for the prediction of tumor growth on experimental data of human and rat origin

Spyridon Patmanidis, Roberto Chignola, Alexandros C. Charalampidis,
George P. Papavassilopoulos

► To cite this version:

Spyridon Patmanidis, Roberto Chignola, Alexandros C. Charalampidis, George P. Papavassilopoulos. A comparison between Nonlinear Least Squares and Maximum Likelihood estimation for the prediction of tumor growth on experimental data of human and rat origin. *Biomedical Signal Processing and Control*, 2019, 54, pp.101639. 10.1016/j.bspc.2019.101639 . hal-02283074

HAL Id: hal-02283074

<https://hal.science/hal-02283074>

Submitted on 11 Sep 2019

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.

A Comparison between Nonlinear Least Squares and Maximum Likelihood Estimation for the Prediction of Tumor Growth on Experimental Data of Human and Rat Origin

Spyridon Patmanidis^{a,*}, Roberto Chignola^b, Alexandros C. Charalampidis^{c,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 Biotechnology, University of Verona, Cà Vignal 1, Strada Le Grazie 15, 37134, Verona, Italy*

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

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

Abstract

Several mathematical models have been developed to explain the growth of tumors and used to fit experimental or clinical data. Their predictive power – i.e. their ability to forecast the future growth on the basis of present knowledge – however, has been rarely explored. Here, we investigate whether a Hidden Markov Model (HMM) based on the well-established Gompertz tumor growth function with additive Gaussian noise could effectively be used to predict the future growth of experimental tumors. The idea behind this work is that one might achieve more accurate predictions if estimates of the unknown parameters of the HMM are used instead of those obtained by fits of the deterministic Gompertz model to the data. We use the principle of Maximum Likelihood (ML) to estimate unknown parameters related to growth dynamics and noise,

*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),
roberto.chignola@univr.it (Roberto Chignola),
alexandros.charalampidis@centralesupelec.fr (Alexandros C. Charalampidis),
yorgos@netmode.ece.ntua.gr (George P. Papavassilopoulos)

and we compare its effectiveness to the classical Nonlinear Least Squares minimization approach. The analyses show that our approach can provide better growth predictions when the data contain adequate information concerning the tumors saturation phase. The forecasts could also be improved by taking into account prior knowledge about the unknown parameters when the information concerning the saturation phase was inadequate. We conclude that by using HMMs in combination with the principle of ML, one can obtain more reliable growth predictions for individual tumors.

Keywords: Tumor Growth, Nonlinear Systems, Parameter Estimation, Maximum Likelihood, Nonlinear Least Squares, Hidden Markov Model, Noise.

1. Introduction

Human cancers are usually treated with one or a combination of therapies such as chemotherapy, immunotherapy, radiotherapy, and surgery. In order to improve the therapeutic outcome, the use of mathematical models along with
5 optimal control has been proposed by many researchers [1–4]. Michor and Beal [5] present extensive literature on the topic of how mathematical modeling can be applied to deliver better drug regimens. In this context, the importance of mathematical modeling is apparent since the models’ ability to describe, and eventually predict, tumor growth greatly affects the outcome of the therapy.
10 However, there is no general consensus about which could be the best mathematical model of tumor growth.

Part of the problem is the complexity of the tumor biology. For example, in the case of solid tumors cancer and normal cells actively communicate by means of a number of molecular signals and conspire to shape the tumor microenvi-
15 ronment and finally to regulate the overall growth of the tumor itself. The molecular network includes positive and negative feedbacks and its nonlinear character challenges mathematical modeling. Models that attempt to capture this microscopic complexity have been developed, but their use in the clinical settings is limited owed to their high computational costs [6]. For this reason,

20 simple phenomenological models developed in the past are still used to describe
tumor growth. The limited number of observations that can be collected in
the clinical settings, however, does not allow to obtain enough data to compare
the accuracy of different growth models. For example, whether the growth of
real tumors is limited above or not is still a matter of debate [7]. As a further
25 complication, it has been recently shown that even in the case of blood tumors,
where it is much easier to draw tumor cells from patients and study their growth
kinetics, some clones appear to grow exponentially while others show sigmoidal
growth up to a steady-state level [8].

Experimental solid tumors can provide data to test mathematical models,
30 and in this case it has been shown that the Gompertz model is a good analytical
descriptor of the macroscopic features of tumor growth and of biological growth
in general [9–13]. It is a phenomenological model that describes exponential
growth with an exponentially decreasing growth rate, and it has also been shown
to approximate quite reasonably the behavior of biologically-motivated, but
35 more complex, tumor growth models [14]. Its mathematical simplicity but at the
same time its ability to capture the macroscopic features of experimental tumors
make the Gompertz model a good choice to test new forecasting approaches.

In our previous works, we proposed and tested with synthetic data a method-
ology that estimates the unknown parameters of the Gompertz model [15–17].
40 The first one is called growth rate and is related to the cells ability to proliferate,
while the second one is called carrying capacity and is related to the tumors sat-
uration volume. In the literature based to the findings of Brunton and Wheldon
[11–13], the carrying capacity is usually considered as a fixed species-specific
parameter when the Gompertz model is used to fit curves to experimental data
45 [18–20], and a Nonlinear Least Squares estimator is used to estimate the growth
rate. This approach usually provides curves that fit well to the data, espe-
cially when the number of available measurements is large and the tumors have
reached a growth maturity. Yet, it ignores the observed heterogeneity of tumor
growth that has been recognized to have clinical implications in the planning
50 of effective treatment schedules [21–24]. This heterogeneity is likely to stem

from the peculiar microscopic characteristics of each tumor, features that at the macroscopic level translate into growth behaviors, and thus Gompertz growth parameters, that vary between individual tumors [21–24].

Taking that heterogeneity into consideration, we proposed a Hidden Markov
55 Model (HMM) based on the Gompertz function that considers both the carrying capacity and the growth rate as patient–specified parameters. In order to estimate the unknown parameters of the HMM, we used the principle of Maximum Likelihood [25]. The properties of Maximum Likelihood Estimators (MLEs) have been extensively in the past and can be also found in [25]. The
60 MLE implementation was presented in detail in [17] and was able to estimate the growth parameters of the Gompertz model with accuracy on synthetic tumor data. The difficulty of working with HMMs lies in the inability to measure the hidden states. However, the proposed MLE algorithm can successfully deal with this challenge and provide accurate estimates of the unknown parameters.
65 Some other challenges are the uncertainty which arises due to the small size of the datasets and the calibration that the algorithm needs in order to provide accurate estimates [16, 17].

In this work, experimental data from past works are being used [14, 19, 26, 27]. The data are time series of volume size measured for individual multicell
70 tumor spheroids (MTSs) *in vitro*. MTSs are three–dimensional aggregates of tumor cells that can be obtained and cultured *in vitro* under controlled experimental conditions [28]. They represent a tumor model with an intermediate complexity between standard two–dimensional monolayer cultures *in vitro* and *in vivo* tumors, as they approximate many biological characteristics of small
75 non–vascularized tumors or of intervascular regions of larger tumors [28]. These include – but are not limited to – the heterogeneous expression of cell surface markers, the production of an intercellular matrix, the heterogeneous distribution of nutrients and of waste molecules and the presence of a central core of quiescent and eventually dead cells [28]. MTSs show a neat spherical morphology and this fact is commonly exploited to convert their diameters, measured at
80 the microscope or by image analysis, into spheroid volumes [14, 26–28]. In

turn, this also means that we can summarize the microscopic three-dimensional complexity of the spheroids in just one measured parameter and thus use a simple one-dimensional model to describe their growth kinetics. In particular, the growth kinetics of MTSs are well described by the Gompertz model [14, 24, 26, 27], and thus MTSs are an excellent experimental system to test the accuracy of new forecasting methods based on the Gompertz growth law.

The data have been obtained with different cell types from human and rat tumors, and thus they can allow to take into account the species-specific character of the Gompertz model discussed above. In particular we analyze the time-dependent volume growth of individual spheroids obtained with human breast carcinoma (MCF7) and glioblastoma – a brain cancer – (U118) cells and with rat glioblastoma tumors (9L) [19]. Along with the MLE from our previous works and the commonly used NLS approach, we implemented two variations for the NLS approach. In the first one, we consider the carrying capacity as an unknown parameter to be estimated. In the second one, instead of fixing the carrying capacity to the maximum attained value (MAV) from the available dataset, we fixed the carrying capacity to the mean value of the mean of the last n measurements of each dataset (mentioned as Mean of Means – MM). This work aims to answer two main questions: 1. whether the proposed model can be used effectively to describe and predict the tumor growth of experimental data and 2. how this approach performs if compared to the standard nonlinear fit of the Gompertz model to growth data.

In the following sections, we present the materials and describe the methods (Section 2), present the simulation results (Section 3), explain our findings (Section 4) and discuss future work (Section 5).

2. Materials and Methods

2.1. Data

The MTSs were obtained using U118 (human glioblastoma), 9L (rat glioblastoma) [19, 26, 27] and MCF7 (human breast carcinoma) cell lines [14]. Overall,

32 9L spheroids, 8 U118 spheroids, and 5 MCF7 spheroids were considered for growth analysis. A detailed description about how the cancer cells were cultured to obtain spheroids can be found in the aforementioned references [14, 19, 26, 27]. Some differences between these data sets, however, deserve to
115 be discussed. U118 and 9L spheroids were isolated by micromanipulation from bulk cultures when their size reached ~ 200 μm . Then the growth of individual spheroids was monitored for ~ 60 days (9L MTSs) or ~ 30 days (U118 MTs) by measuring spheroid size with an inverted microscope equipped with a calibrated ocular micrometer. On average the 9L dataset comprises 64.6 ± 9.7
120 volume measurements for each spheroid (min. 47, max. 79) whereas the U118 dataset 32.7 ± 3.2 volume measurements/spheroid (min. 28, max. 35). U118 spheroids were indeed difficult to obtain and maintain in culture for long times and this explain why this set include a lower number of spheroids measured for a shorter time span if compared to the 9L dataset.

125 The experimental conditions used to obtain 9L and U118 spheroids allowed in particular to investigate the final plateau phase of the sigmoidal growth kinetics, whereas the first growth phase could not be observed. MCFT spheroids were therefore obtained from cloned cells and grown for ~ 60 days. The cells were cloned by seeding them at the average limiting dilution of 0.1 cells/well
130 into the wells of five 96-well culture plate. Limiting dilution experiments obey Poisson statistics [29] and under these conditions one may expect to obtain a total of ~ 43 wells containing exactly 1 cell. However, *in vitro* the plating efficiency – i.e. the ability of single cells to form a clonal population – is never 100% and actually it varies between 20–30%. Indeed, 11 wells containing 1
135 cell were obtained and only 5 clones could form spheroids. The size of MCF7 spheroids was measured from calibrated micrographs taken with a digital microscope (EVOsci, AMG, Bothell, WA, USA) by using the image analysis software ImageJ (<http://imagej.nih.gov>). Starting from single cells a time period of ~ 60 days was appropriate to accurately determine the first quasi-exponential growth
140 phase of individual spheroids but not to characterize the plateau phase of their growth kinetics. On average the MCF7 dataset comprises 24.8 ± 8.6 volume

measurements/spheroid (min. 15, max. 33).

2.2. Tumor Growth Modeling

The stochastic discrete time state–space representation of the Gompertz model, which was introduced in [16] and also used in [17], was used here to describe the growth dynamics. This model can be formulated as follows:

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

$$y_k = x_k + v_k, \quad (1b)$$

where x_k (mm^3) is the tumor volume at time step k , y_k the a measurement that corresponds to the state x_k , and $k \in \mathbb{N}$. The function f in Eq. (1a) is the state–space representation of the Gompertz function [30]:

$$f(x_k, \theta_1, \theta_2) = \theta_2 \exp \left(\ln \left(\frac{x_k}{\theta_2} \right) \exp \left(- \frac{1}{\theta_1} t_{k+1} \right) \right), \quad (2)$$

where θ_1 (days) is related to the cells' proliferative ability, θ_2 (mm^3) is the carrying capacity ($\lim_{k \rightarrow \infty} x_k = \theta_2$) and t_{k+1} (days) is the time between k and $k + 1$. The random variables w_k and v_k , $k \in \mathbb{N}$ are mutually independent and normally distributed with zero mean and unknown variance: $w_k \sim \mathcal{N}(0, \sigma_{w_k}^2)$ and $v_k \sim \mathcal{N}(0, \sigma_{v_k}^2)$, where $\sigma_{w_k} = \theta_3 x_k^{\theta_4}$ and $\sigma_{v_k} = \theta_5 x_k^{\theta_6}$. Measurements are available from time $k=1$ onwards. Because of the random components w_k and v_k , the model of Eq. (1) also corresponds to the description:

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

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

where $p_{\theta_a}(x_{k+1}|x_k)$ is the probability density function describing the dynamics
145 for given values of x_k and $p_{\theta_b}(y_k|x_k)$ is the probability density function describing the measurements [31], [32]. 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 $\Theta_b \subseteq \mathbb{R}_{>0}^2$ denoting a compact set of permissible values of the unknown vector θ_b . We also
150 define the vector $\theta = [\theta_a, \theta_b]$, where $\theta \in \Theta$ with $\Theta \subseteq \mathbb{R}_{>0}^6$.

2.3. Parameter Estimation

The problem addressed in this manuscript is how to obtain an estimate $\hat{\boldsymbol{\theta}}$ based on N measurements $\mathbf{Y}_N = [y_1, \dots, y_N]$ and predict the future states $\hat{\mathbf{X}}_{N+1:M} = [\hat{x}_{N+1}, \dots, \hat{x}_M]$. In order to estimate the unknown parameters implemented methods based on the Nonlinear Least Squares and the Maximum Likelihood. In this manuscript, we include a brief description of the Maximum Likelihood estimator.

2.3.1. Nonlinear Least Squares Estimator

This method is used to find the estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ of the parameters θ_1 and θ_2 that minimize the summation of the squared difference between the measurements \mathbf{Y}_n and the estimates $\hat{\mathbf{X}}_N$:

$$[\hat{\theta}_1, \hat{\theta}_2] = \arg \min_{\theta_1 \in \Theta_1, \theta_2 \in \Theta_2} \sum_1^N (y_k - \hat{x}_k)^2. \quad (4)$$

where $\theta_1 \in \Theta_1$ with $\Theta_1 \subseteq \mathbb{R}_{>0}$ and $\theta_2 \in \Theta_2$ with $\Theta_2 \subseteq \mathbb{R}_{>0}$ denote compact sets of permissible values for the unknown parameters θ_1 and θ_2 , and x_k is given by:

$$\hat{x}_k = \theta_2 \exp \left(\ln \left(\frac{\hat{x}_{k-1}}{\theta_2} \right) \exp \left(- \frac{1}{\theta_1} t_k \right) \right). \quad (5)$$

The modifications needed in order to test the cases in which the carrying capacity is fixed to the MAV or MM are trivial.

2.3.2. Maximum Likelihood Estimator

This method is applied to the model of Eq. (1) 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:

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta} \in \Theta} p_{\boldsymbol{\theta}}(\mathbf{Y}_N). \quad (6)$$

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

$$p_{\boldsymbol{\theta}}(\mathbf{Y}_N) = p_{\boldsymbol{\theta}}(y_1) \prod_{k=2}^N p_{\boldsymbol{\theta}}(y_k | \mathbf{Y}_{k-1}), \quad (7)$$

where

$$p_{\boldsymbol{\theta}}(y_{k+1} | \mathbf{Y}_k) = \int p_{\boldsymbol{\theta}}(y_{k+1} | x_{k+1}) p_{\boldsymbol{\theta}}(x_{k+1} | \mathbf{Y}_k) dx_{k+1}, \quad (8)$$

$$p_{\theta}(x_{k+1}|\mathbf{Y}_k) = \int p_{\theta}(x_{k+1}|x_k)p_{\theta}(x_k|\mathbf{Y}_k)dx_k, \quad (9)$$

and

$$p_{\theta}(x_{k+1}|\mathbf{Y}_{k+1}) = \frac{p_{\theta}(y_{k+1}|x_{k+1})p_{\theta}(x_{k+1}|\mathbf{Y}_k)}{p_{\theta}(y_{k+1}|\mathbf{Y}_k)}. \quad (10)$$

The distributions in Eqs. (8) to (10) can be computed iteratively. However, in most cases, as also in the case studied in this manuscript, the above integrals cannot be solved analytically. Nevertheless, since x_k is low-dimensional, numerical integration can be used effectively in order to compute the integrals of Eqs. (9) and (10). More specifically, we divide the integration interval into subintervals and we compute each subintegral by using the trapezoidal rule. The summation of these subintegrals provides an approximation of the initial integral we want to compute. Moreover, to reduce the interval of integration $[0, X_{\max}]$ the Gaussian property of the noise is being used. At every iteration step, x_k is considered a normally distributed random variable with mean $\mu = y_k$ and variance $\sigma^2 = (\sigma_0 y_k^{e_0})^2$ ($x_k \sim \mathcal{N}(\mu, \sigma^2)$). As a result, the interval $[\mu - 5\sigma, \mu + 5\sigma] \cap [0, X_{max}]$, which is significantly smaller than $[0, X_{max}]$, can be used as the interval of integration [17]. For a more detailed description of the Maximum Likelihood Estimator, the reader may refer to Section 3.1.3 Numerical Maximum Likelihood in [17].

3. Results

In this section, we present the simulation results. In order to perform these simulations we used MATLAB (version R2015a, The MathWorks Inc., Natick, IL). The computational methods were implemented by writing suitable functions, while Eqs. (4) and (6) were computed with the `fmincon` function [33].

In this work, we examined three different datasets that contain multicell tumor spheroids. MCF7 refers to human breast carcinoma cell line, U118 to human glioblastoma cell line and 9L to rat glioblastoma cell line. Fig. (1) shows plots of the growth data for all datasets in order to inform the reader on the type of experimental data we used for the simulations. We implemented

the ML estimator proposed in [17] and a Least Squares estimator and compared their performance. For the ML estimator, both the growth rate and the carrying capacity were estimated at each time step. As regards the LS estimator, along with the approach proposed in the literature where the carrying capacity is fixed to a certain known value, we also examined the case in which this parameter is considered unknown (referred below as LS_{2p}). In the case that the carrying capacity is fixed, two sub-cases were studied. In the first one, the MAV of each dataset was considered as the fixed value of carrying capacity (referred below as LS_{MAV}), while in the second one the MM (referred below as LS_{MM}). The value of MM is given by:

$$MM = \frac{1}{nM} \sum_{i=1}^M \sum_{j=N-n}^N y_{i,j}, \quad (11)$$

where n is the number of measurements used to obtain the mean carrying capacity of the spheroids, M is the number of spheroids the dataset contains, i is the spheroid's ID and j is the measurement's ID. In our simulations, n has been set to 10 for the U118 and the 9L datasets, while for the MCF7 dataset 185 it has been set to 5 due to the small number of available measurements. The parameter estimates were used in order to predict the tumors' future growth.

The Root-Mean-Square Deviation (RMSD) was used in order to measure the accuracy of the predictions. Given $Y_N = [y_1, \dots, y_N]$, the first k measurements ($Y_{1:k} = [y_1, \dots, y_k]$, $k < N$) were used to estimate the growth dynamics $\hat{\theta}_1$ and $\hat{\theta}_2$. The future growth of the tumor ($\hat{X}_{k+1:N} = [\hat{x}_{k+1}, \dots, \hat{x}_N]$) was predicted by the following equation:

$$\hat{x}_i = \hat{\theta}_2 \exp \left(\ln \left(\frac{\hat{x}_{i-1}}{\hat{\theta}_2} \right) \exp \left(- \frac{1}{\hat{\theta}_1} t_i \right) \right), \quad (12)$$

where $i = 2, \dots, N$ and $\hat{x}_1 = y_1$. Then, the RMSD was computed by:

$$RMSD = \sqrt{\frac{\sum_{i=k+1}^N (y_i - \hat{x}_i)^2}{N - (k + 1)}} \quad (13)$$

The RMSD of the LS estimator will be referred to as $RMSD_{LS}$ and the RMSD of the ML estimator as $RMSD_{ML}$.

Table 1: Number of tumors for which the RMSD of the predictions is lower for the Maximum Likelihood approach. The Maximum Likelihood approach is compared to each Least Squares approach separately. This table refers to the MCF7 dataset.

MCF7												
No. of Available Measurements												
	2	3	4	5	6	7	8	9	10	11	12	13
ML vs LS_{MAV}												
No. of Tumors	1	4	2	1	1	1	0	0	0	1	0	0
ML vs LS_{MM}												
No. of Tumors	2	3	4	3	3	2	2	2	2	2	1	1
ML vs LS_{2p}												
No. of Tumors	2	3	3	2	2	2	1	0	1	1	0	0

190 Tables. (1)–(3) show the number of cases at each time step where $RMSD_{ML}$ was lower than $RMSD_{LS}$ (cases where the ML performed better). The ML method is compared to each Least Squares approach for each dataset separately. We also present the comparison between the mean RMSD of the prediction at each time step of ML and LS_{MAV} in Figs. (2), of ML and LS_{MM} in Figs. (3),
 195 and of ML and LS_{2p} in Figs. (4).

4. Discussion

Even though the ability of a broad range of mathematical models (including Gompertz) to describe experimental data is studied in several studies, the predictive power is rarely considered, as Benzekry et al. point in [7]. This
 200 work examines whether an HMM based on the Gompertz function with additive Gaussian noise can be used effectively in order to describe and predict the growth of experimental tumor data. This model entails the challenge of computing the prediction density $p_{\theta}(y_k|Y_{k-1})$ in absence of the hidden states x_k . Our main hypothesis is that we can achieve more accurate predictions for the future
 205 growth of a tumor if we use the estimates of the unknown parameters of the

Table 2: Number of tumors for which the RMSD of the predictions is lower for the Maximum Likelihood approach. The Maximum Likelihood approach is compared to each Least Squares approach separately. This table refers to the U118 dataset.

U118										
No. of Available Measurements										
	2	3	4	5	6	7	8	9	10	11
ML vs LS_{MAV}										
No. of Tumors	5	2	3	3	3	2	2	4	4	5
ML vs LS_{MM}										
No. of Tumors	4	4	4	3	4	3	3	2	7	6
ML vs LS_{2p}										
No. of Tumors	5	7	6	6	3	3	3	5	5	6
No. of Available Measurements										
	12	13	14	15	16	17	18	19	20	
ML vs LS_{MAV}										
No. of Tumors	5	5	5	5	3	5	4	5	3	
ML vs LS_{MM}										
No. of Tumors	4	4	5	5	4	4	4	5	4	
ML vs LS_{2p}										
No. of Tumors	4	4	5	5	3	5	3	3	1	

Table 3: Number of tumors for which the RMSD of the predictions is lower for the Maximum Likelihood approach. The Maximum Likelihood approach is compared to each Least Squares approach separately. This table refers to the 9L dataset.

9L										
No. of Available Measurements										
	2	3	4	5	6	7	8	9	10	11
ML vs LS_{MAV}										
No. of Tumors	31	29	29	32	32	31	30	30	31	30
ML vs LS_{MM}										
No. of Tumors	4	9	7	7	5	5	5	6	9	8
ML vs LS_{2p}										
No. of Tumors	13	16	18	18	16	22	17	17	16	16
No. of Available Measurements										
	12	13	14	15	16	17	18	19	20	21
ML vs LS_{MAV}										
No. of Tumors	27	27	28	26	27	27	27	25	27	29
ML vs LS_{MM}										
No. of Tumors	10	14	16	16	16	15	16	17	22	23
ML vs LS_{2p}										
No. of Tumors	16	16	12	13	12	12	8	13	18	19
No. of Available Measurements										
	22	23	24	25	26	27	28	29	30	
ML vs LS_{MAV}										
No. of Tumors	30	29	29	29	30	31	31	30	32	
ML vs LS_{MM}										
No. of Tumors	25	23	25	21	24	23	24	23	24	
ML vs LS_{2p}										
No. of Tumors	20	14	16	18	23	23	22	19	23	

HMM instead of the estimates obtained by the classical approach of the deterministic Gompertz model with the carrying capacity fixed to a species-specific parameter. We also hypothesize that the growth predictions can be improved by utilizing prior knowledge about the unknown parameters. In order to test
210 these hypotheses, we used MTSs from three different types of tumors.

The novelty of this work is that an HMM with additive Gaussian noise that considers the growth rate, the carrying capacity and the noise characteristics as unknown parameters is used to describe the growth of experimental tumor data. Furthermore, in order to estimate the unknown parameters, we used
215 an MLE algorithm that we developed in our previous works, instead of the general Expectation Maximization algorithm [34] which is usually employed for the computation of the Maximum Likelihood estimates. The advantage of our MLE algorithm is that it exploits the structure of the used HMM in order to provide a very accurate approximation of the prediction density. Moreover, the
220 HMM approach which has been used in this work for tumor growth modeling, can be considered as a general method to forecast biological time series, since the Gompertz function can be replaced by any nonlinear equation that models a biological procedure. If, for example, the Gompertz function is replaced by a model that fits data of pathogens' spreading in humans or in plants, this
225 approach can be used to forecast the kinetics of pathogens' spreading. Thus, the proposed method can be used not only for cancer research, but also for parameter estimation and forecasting in epidemiology [35–37].

As regards the MCF7 dataset, all the NLS estimators performed better than MLE. This could be an indication that the stochastic Gompertz model which
230 has been used is not suitable for describing this kind of tumors. However, a closer look at Figs. (1a) reveals that until the 13th measurement (which is the time window when we try to estimate the growth rate and the carrying capacity in order to predict the future growth) the spheroids are far from the saturation phase, which means that there is not enough information considering
235 the carrying capacity. Indeed this experimental set was obtained with cloned cells, i.e. the data are from tumor spheroids that started their growth when they

were formed by one cell only, with the aim to investigate in details the initial quasi-exponential growth phase. The plateau phase is not well characterized (see also Fig 1) and this is a critical point since the ML estimates the carrying capacity directly from the data. Further research with a larger dataset that
240 monitors the growth for a longer time interval data could lead to better results.

In regard to the U118 dataset, we can see a different behavior. The NLS approach with the carrying capacity fixed to the MAV performs better than MLE at the early growth stages (when there are only a few available measurements).
245 As the growth progressed, the MLE approach provides more accurate predictions in the majority of cases, even though the mean RMSD_{LS} is slightly lower than the mean RMSD_{ML} . This behavior is expected since the saturation phase for this dataset has been reached during the examined time window. When the carrying capacity is unknown, the performance of NLS drops and MLE performs
250 generally better for both the early and later growth stages. Lastly, using the MM as carrying capacity has a positive influence at the early and negative at later growth stages.

9L spheroids were monitored for a longer period of time (thus there are more samples for each spheroid, see Section 2.1) and the saturation phase was reached
255 more quickly. As a result, for this dataset, the MLE over-performs the NLS approach. By considering the carrying capacity unknown and to be estimated, the performance of the NLS estimator improved, but the ML estimator still performed slightly better. However, using the MM as carrying capacity greatly improved the NLS prediction ability at early stages. At later growth stages, the
260 MLE is again able to provide more accurate predictions.

Overall, our analyses indicate that the monitoring time interval is a decisive factor for the prediction of the future growth. Whether or not the data contain enough information for the saturation phase is therefore critical to make good predictions of tumor growth with the Gompertz model. When the data lack
265 this information, the approaches that fix the carrying capacity are expected to have better performance since the fixed parameter includes some information about the saturation phase. However, even though the approaches with the fixed

parameter may provide curves that fit well to the data, they are in many cases unable to predict the tumors future growth, because the information about the saturation phase might be inaccurate (due to the intrinsic growth variability of individual tumor spheroids even when they are produced with cells from the same cell line [24]).

As regards our first hypothesis, we confirmed that the proposed HMM is able to describe the growth of experimental tumors. Furthermore, we observed that when a tumor starts to grow rapidly and the measurements contain information about the saturation phase, this approach provided, in most cases, more accurate predictions of future growth. This is however a potential problem since, as far as we know, there are no firm data concerning the saturation phase of *in vivo* solid tumors because in general they are treated a short time after their diagnosis. It is tempting to speculate that each organ can nonetheless contain tumors of a given maximum size and it might be interesting to infer the carrying capacity of different organs on the basis of some biophysical characteristics, like e.g. their biomechanical properties, or from statistical analyses of clinical cancer data such as tumor size at patients' death. For example, it is well known that solid-stress from the surrounding environment can inhibit the growth of experimental tumors and affect the plateau phase of their growth kinetics [38]. Tissue-specific estimates might then be used as prior knowledge to inform and possibly validate our approach also for *in vivo* tumors. Indeed, we show that the utilization of prior knowledge in the NLS estimator can improve the growth predictions at early growth stages, but as the tumor grows, we can obtain estimates that are able to predict tumor growth more accurately by combining the HMM model with the MLE.

At this point, we would like to notice some ideas for further research that might be of interest. First of all, it would be very interesting to examine the descriptive power of models other than the Gompertz. To put it in another way, instead of the stochastic model of Eq. (1), the MLE can be combined with an HMM (based on another growth law) that can describe certain cancer tumors better than the Gompertz. This could lead to a more accurate growth

prediction for particular types of tumors. Another interesting topic for further
300 research is to examine noise distributions other than the Gaussian. In our works,
we chose to work with Gaussian distributed noise, because it is probably the
most common choice when there is no information available concerning the noise
distribution. However, even though we obtained good results by using Gaussian
distributed noise, there is no evidence that this is the best choice to model the
305 uncertainties for the problem we studied. Lastly, since there is evidence that
prior information can improve the growth predictions, using estimators that
take prior knowledge about the unknown parameters into consideration has the
potential to provide more accurate results. For this purpose, we believe that
examining the performance of the Maximum A Posteriori estimator would be
310 very interesting for future research, since this estimator combines the principle of
Maximum Likelihood with the prior knowledge about the unknown parameters.

5. Conclusions

In our previous works, we proposed a new approach for the estimation of the
Gompertz's model parameters and tested it on synthetic tumor data. In this
315 work, we applied that methodology on experimental tumor data and evaluated
its potential to predict the future growth of a tumor.

The main contribution is that this work shows the ability of the proposed
HMM to describe the tumor growth of experimental data. Furthermore, it indi-
cates that the proposed approach can be used for tumor growth prediction and
320 that when the measured data contain enough information about the saturation
phase of the tumor, it can provide more accurate tumor growth predictions com-
pared to the classical approach. Lastly, it shows that the utilization of the prior
knowledge concerning the unknown parameters has the potential to improve the
growth predictions at early growth stages where the data may lack information
325 about the tumors saturation phase.

To conclude, it should be noted that our estimation technique is not limited
to the Gompertz model, and thus the MLE can be applied to other models

that may describe tumor growth models more appropriately (e.g. [14]). We believe that individualized tumor growth modeling is of great importance since
330 it takes into consideration the heterogeneous growth of individual tumors and thus it has the potential to provide more accurate results. The MLE is a general approach that can be exploited for this purpose.

Conflict of interest

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

Acknowledgment

Spyridon Patmanidis wishes to thank the Special Account for Research of the National Technical University of Athens for supporting his doctoral research
340 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 Skłodowska-Curie grant agreement No 705982.

References

- 345 [1] 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.
- [2] 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.
350

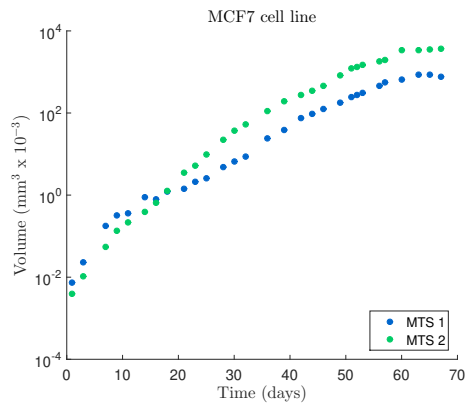
- [3] R. Serre, S. Benzekry, L. Padovani, C. Meille, N. André, J. Ciccolini, F. Barlesi, X. Muracciole, D. Barbolosi, Mathematical modeling of cancer immunotherapy and its synergy with radiotherapy, *Cancer Research* 76 (17) (2016) 4931–4940. doi:10.1158/0008-5472.CAN-15-3567. 355
- [4] S. Benzekry, A. Tracz, M. Matri, 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.
- [5] F. Michor, K. Beal, Improving cancer treatment via mathematical modeling: Surmounting the challenges is worth the effort, *Cell* 163 (5) (2015) 1059–1063. doi:http://dx.doi.org/10.1016/j.cell.2015.11.002. 360
- [6] J. Metzcar, Y. Wang, R. Heiland, P. Macklin, A review of cell-based computational modeling in cancer biology, *JCO Clinical Cancer Informatics* 3 (2019) 1–13. doi:10.1200/CCI.18.00069. 365
- [7] S. Benzekry, C. Lamont, A. Beheshti, A. Tracz, J. M. L. Ebos, L. Hlatky, P. Hahnfeldt, Classical mathematical models for description and prediction of experimental tumor growth, *PLOS Computational Biology* 10 (8) (2014) 1–19. doi:10.1371/journal.pcbi.1003800.
- [8] M. Gruber, I. Bozic, I. Leschiner, D. Livitz, K. Stevenson, L. Rassenti, D. Rosebrock, A. Taylor-Wiener, O. Olive, R. Goyetche, S. M. Fernandes, J. Sun, C. Stewart, C. C. A. Wong, W. Zhang, J. G. Reiter, J. M. Gerold, J. G. Gribben, K. R. Rai, M. J. Keating, J. R. Brown, D. Neuberg, T. J. Kipps, M. A. Nowak, G. Getz, C. J. Wu, Growth dynamics in naturally progressing chronic lymphocytic leukemia, *Nature* (2019) doi:https://doi.org/10.1038/s41586-019-1252-x. 370 375
- [9] A. K. Laird, Dynamics of tumor growth, *British Journal of Cancer* 18 (1964) 490–502.

- [10] A. K. Laird, Dynamics of growth in tumors and normal organisms, National
380 Cancer Institute Monograph 30 (1969) 15–28.
- [11] G. F. Brunton, T. E. Wheldon, Prediction of the complete growth pat-
tern of human multiple myeloma from restricted initial measurements, *Cell
Proliferation* 10 (6) (1977) 591–594. doi:10.1111/j.1365-2184.1977.
tb00316.x.
- 385 [12] 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.
- [13] G. F. Brunton, T. E. Wheldon, The gompertz equation and the construc-
tion of tumour growth curves, *Cell Proliferation* 13 (4) (1980) 455–460.
390 doi:10.1111/j.1365-2184.1980.tb00486.x.
- [14] E. Milotti, V. Vladislav, M. Sega, R. Chignola, Interplay between distribu-
tion of live cells and growth dynamics of solid tumours, *Scientific Reports*
2 (990) (2012) 1–7. doi:http://dx.doi.org/10.1038/srep00990.
- [15] B. Gompertz, On the nature of the function expressive of the law of human
395 mortality, and on a new mode of determining the value of life contingencies,
Philosophical Transactions of the Royal Society of London 115 (1825) 513–
583.
- [16] S. Patmanidis, A. C. Charalampidis, I. Kordonis, G. D. Mitsis, G. P. Pa-
pavassilopoulos, Comparing methods for parameter estimation of the gomp-
400 pertz tumor growth model, *IFAC-PapersOnLine* 50 (1) (2017) 12203–12209.
doi:https://doi.org/10.1016/j.ifacol.2017.08.2289.
- [17] S. Patmanidis, A. C. Charalampidis, I. Kordonis, G. D. Mitsis, G. P. Pa-
pavassilopoulos, Tumor growth modeling: Parameter estimation with maxi-
mum likelihood methods, *Computer Methods and Programs in Biomedicine*
405 160 (2018) 1–10. doi:https://doi.org/10.1016/j.cmpb.2018.03.014.

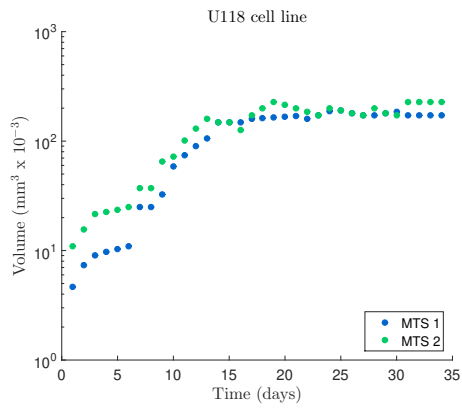
- [18] C. Loizides, D. Iacovides, M. M. Hadjiandreou, G. Rizki, A. Achilleos, K. Strati, G. D. Mitsis, Model-based tumor growth dynamics and therapy response in a mouse model of de novo carcinogenesis, *PLOS ONE* 10 (12) (2015) 1–18. doi:10.1371/journal.pone.0143840.
- 410 [19] R. Chignola, R. I. Foroni, 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.
- [20] 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.
- 415 [21] 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.org/10.1038/264542a0.
- [22] L. Norton, Implications of kinetic heterogeneity in clinical oncology, *Seminars in Oncology* 12 (13) (1985) 231–249. doi:10.5555/uri:pii:009377548590017X.
- 420 [23] H. Bartelink, A. Begg, J. C. Martin, M. van Dijk, L. van 't Veer, P. van der Vaart, M. Verheij, Towards prediction and modulation of treatment response, *Radiotherapy and Oncology* 50 (1) (1999) 1–11. doi:https://doi.org/10.1016/S0167-8140(99)00009-2.
- [24] R. Chignola, R. Foroni, A. Franceschi, M. Pasti, C. Candiani, C. Anselmi, G. Fracasso, G. Tridente, M. Colombatti, Heterogeneous response of individual multicellular tumour spheroids to immunotoxins and ricin toxin, *British Journal of Cancer* 72 (1995) 607–614.
- 430 [25] S. Kay, *Fundamentals Of Statistical Signal Processing*, Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 1993.

- [26] R. Chignola, A. Schenetti, E. Chiesa, R. Foroni, S. Sartpris, A. Brendolan, G. Tridente, G. Andrighetto, D. Liberati, Oscillating growth patterns of multicellular tumour spheroids, *Cell Proliferation* 32 (1) (1999) 39–48. doi: 10.1046/j.1365-2184.1999.3210039.x.
- [27] R. Chignola, A. Schenetti, G. Andrighetto, E. Chiesa, R. Foroni, S. Sartoris, G. Tridente, D. Liberati, Forecasting the growth of multicell tumour spheroids: implications for the dynamic growth of solid tumours, *Cell Proliferation* 33 (4) (2000) 219–229. doi:10.1046/j.1365-2184.2000.00174.x.
- [28] F. Hirschhaeuser, H. Menne, C. Dittfeld, J. West, W. Mueller-Klieser, L. A. Kunz-Schughart, Multicellular tumor spheroids: an underestimated tool is catching up again, *Journal of Biotechnology* 148 (1) (2010) 3–15. doi: <https://doi.org/10.1016/j.jbiotec.2010.01.012>.
- [29] I. Lefkovits, H. Waldmann, Limiting dilution analysis of cells in the immune system, Cambridge University Press, Cambridge, 1979.
- [30] B. Dennis, J. M. Ponciano, Density-dependent state-space model for population-abundance data with unequal time intervals, *Ecology* 95 (8) (2014) 2069–2076. doi:10.1890/13-1486.1.
- [31] T. B. Schön, A. Wills, B. Ninness, System identification of nonlinear state-space models, *Automatica* 47 (1) (2011) 39–49. doi:<https://doi.org/10.1016/j.automatica.2010.10.013>.
- [32] A. C. Charalampidis, G. P. Papavassilopoulos, Development and numerical investigation of new non-linear kalman filter variants, *IET Control Theory Applications* 5 (10) (2011) 1155–1166. doi:10.1049/iet-cta.2010.0553.
- [33] Matlab Optimization Toolbox User’s Guide, The MathWorks Inc., Natick, MA, 2017.

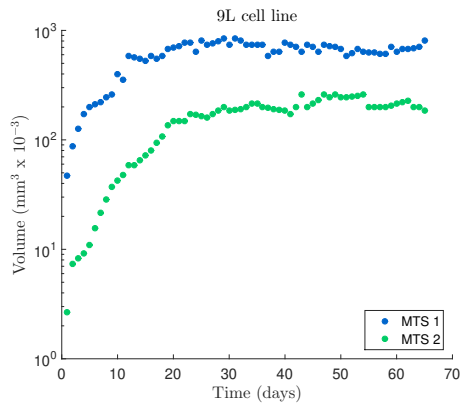
- [34] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from
460 incomplete data via the em algorithm, *JOURNAL OF THE ROYAL STA-
TISTICAL SOCIETY, SERIES B* 39 (1) (1977) 1–38.
- [35] A. Smirnova, G. Chowell, A primer on stable parameter estimation and
forecasting in epidemiology by a problem-oriented regularized least squares
algorithm, *Infectious Disease Modelling* 2 (2) (2017) 268–275. doi:<https://doi.org/10.1016/j.idm.2017.05.004>.
465
- [36] A. Kleczkowski, C. A. Gilligan, Parameter estimation and prediction for
the course of a single epidemic outbreak of a plant disease, *J R Soc Interface*
4 (16) (2007) 865–877. doi:[10.1098/rsif.2007.1036](https://doi.org/10.1098/rsif.2007.1036).
- [37] G. Chowell, P. Fenimore, M. Castillo-Garsow, C. Castillo-Chavez, Sars
470 outbreaks in ontario, hong kong and singapore: the role of diagnosis and
isolation as a control mechanism, *Journal of Theoretical Biology* 224 (1)
(2003) 1–8. doi:[https://doi.org/10.1016/S0022-5193\(03\)00228-5](https://doi.org/10.1016/S0022-5193(03)00228-5).
- [38] G. Helmlinger, P. A. Netti, H. C. Lichtenbled, R. J. Melder, R. K. Jain,
Solid stress inhibits the growth of multicellular tumor spheroids, *Nature*
475 *Biotechnology* 15 (1997) 778–783.



(a) Tumor growth for MCF7 MTS



(b) Tumor growth for U118 MTS



(c) Tumor growth for 9L MTS

Fig. 1. Growth trajectories of multicell tumor spheroids. Representative data on the individual growth of two MCF7, two U118 and two 9L MTS.

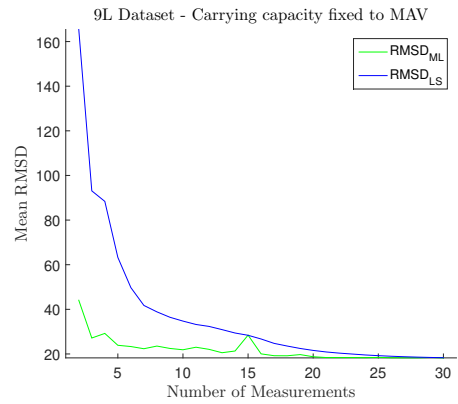
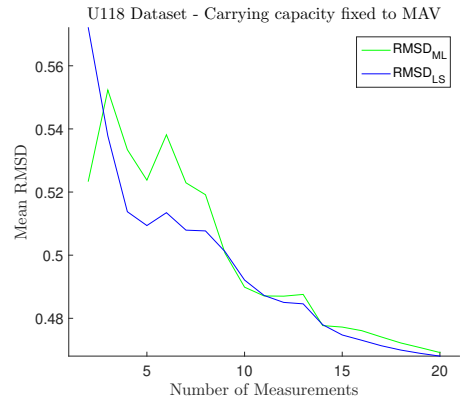
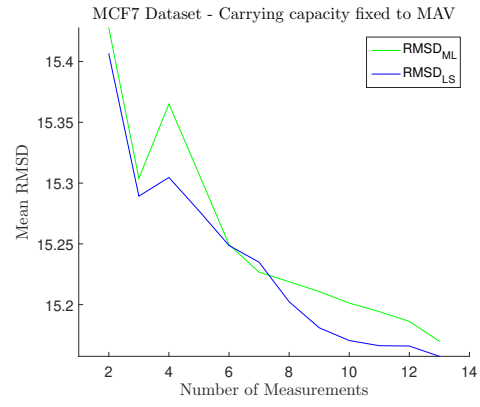


Fig. 2. Mean RMSD for MLE vs NLS_{MAV} .

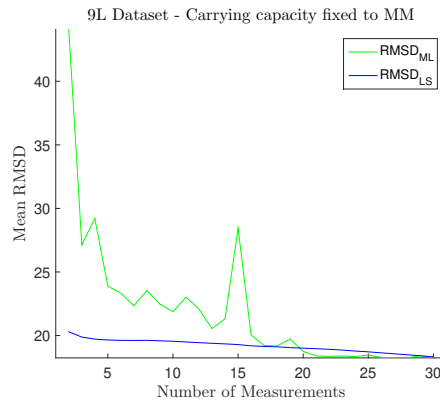
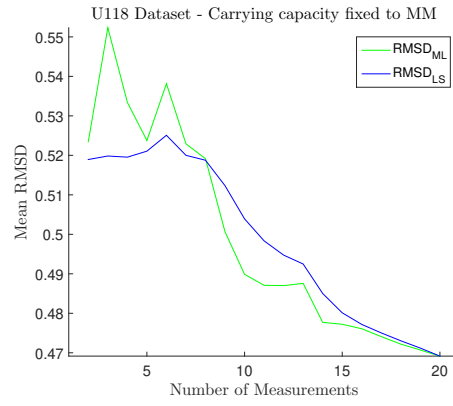
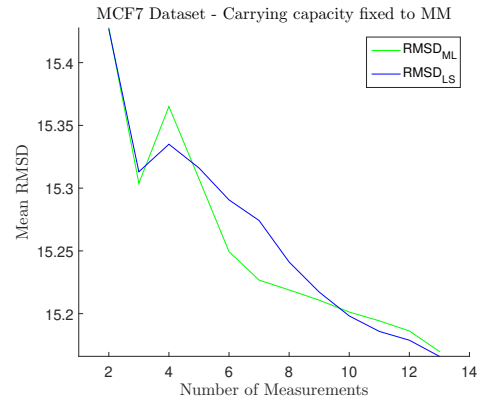


Fig. 3. Mean RMSD for MLE vs NLS_{MM}.

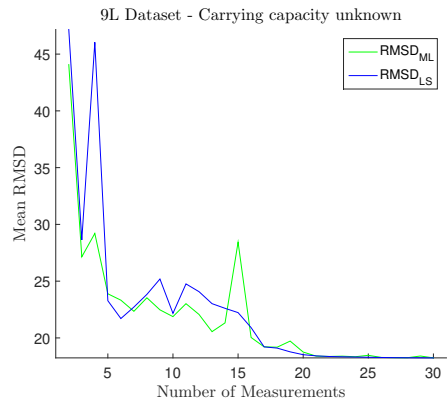
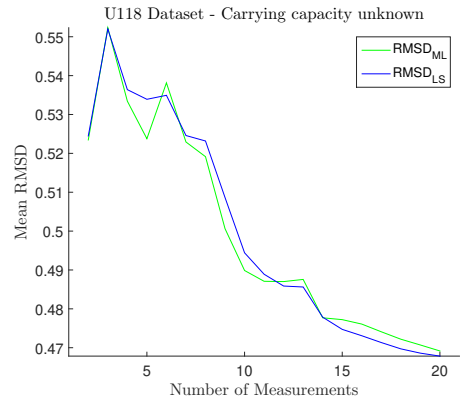
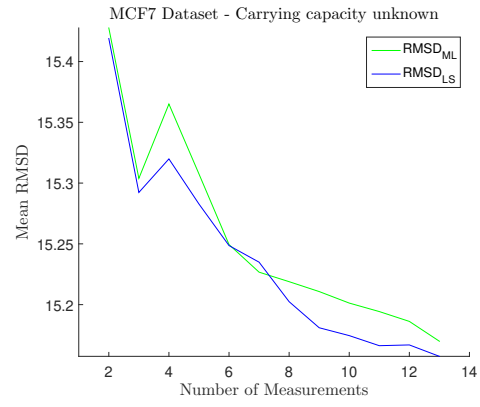


Fig. 4. Mean RMSD for MLE vs NLS_{2p}.