Classical Mathematical Models for Description and Forecast of Preclinical Tumor Growth

Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Lynn Haltky, Philip Hahnfeldt

Classical Mathematical Models for Description and Forecast of Preclinical Tumor Growth

Sébastien Benzekry[a,b], Clare Lamont[a], Afshin Beheshti[a], Lynn Hlatky[a], Philip Hahnfeldt[a]

[a]*Center of Cancer Systems Biology, GRI, Tufts University School of Medicine, Boston, MA, USA*

[b]*Inria Bordeaux Sud-Ouest, Institut de Mathématiques de Bordeaux, Bordeaux, France*

Corresponding author: S. Benzekry, sebastien.benzekry@inria.fr

Word count:

Abstract: 266 (300)

Author summary: 193 (200)

Body text: 6240

Abstract

Tumor growth is a complex process involving a large number of biological phenomena. However, at the macroscopic scale, it seems to follow relatively simple laws that have been formalized with the help of mathematical models.

Based on experimental data of *in vivo* syngeneic tumor growth, rigorous quantitative and discriminant analysis of a wide array of these models was performed for description and prediction of tumor kinetics. Detailed analysis of the measurement error was performed that resulted in the design of an adapted statistical model of the variance error quantifying the uncertainty of the data. Combined to several goodness-of-fit criteria and to a study of the numerical identifiability of the models, it allowed quantification of the descriptive power of each model, from which we inferred insights on macroscopic tumor growth laws. Our analysis enlightens one model as particularly relevant, namely the power growth model, which suggests a novel, simple and minimal theory of neoplastic development based on a fractal dimension of the proliferative tissue.

Detailed study of the predictive properties of the models reveals variable forecasting power among them and quantifies how far and how precise predictions can be made, based on a given number of data points. In situations where small number of data points is available, we studied the effect of adjunction of *a priori* information on the statistical distribution of the parameters during the fit procedure. This method revealed very powerful, yielding significant improvement of the forecast performances, for instance from a 14.9% to a 60.2% success rate when predicting future growth based only on three data points and using the power growth model.

Author Summary

Although depending on a wide array of intricate phenomena, tumor growth results, at the macroscopic scale, in relatively simple time curves that can be quantified using mathematical models. Here we assessed the descriptive and predictive power of the most classical of these in order to infer general laws for the global behavior of neoplastic growth. As a result from our analysis one of the models, namely the power growth model, appears as particularly adapted and proposes a novel theory of tumor growth based on a fractal dimension of the proliferative tissue.

We also assessed the predictive power of these classical mathematical models and show that despite similar descriptive accuracy, models can significantly differ in their ability to predict future growth. When only few data points are available, we propose to use some *a priori* information during the estimation process, a method that reveals very helpful and significantly improves the prediction success rate.

These results could be of value for preclinical cancer research by suggesting what model is best adapted when assessing anti-cancer drugs efficacies. They also offer clinical perspective on what can be expected from mathematical modeling in terms of future growth prediction.

Introduction

Neoplastic growth involves a large number of complex biological processes, including control of the cell cycle, stroma recruitment, angiogenesis or escape from immune surveillance, that in combination produce a macroscopic expansion, thus raising the prospect of a possible general law for the global dynamics of neoplasm. Quantitative and qualitative aspects of the temporal development of tumor growth can be studied in a variety of experimental settings, including *in vitro* proliferation assays, three-dimensional *in vitro* spheroids, syngeneic or xenograft *in vivo* implants (subcutaneous or at the orthotopic site), transgenic mice models and longitudinal studies of clinical images. Each scale has its own advantages and drawbacks, with increasing relevance tending to coincide with decreasing measurement precision. Here we focus on a model of syngeneic Lewis Lung carcinoma (LLC) subcutaneously implanted in immune-competent mice.

Global quantification of tumor cell kinetics is a full discipline, termed cytokinetics [1], that can be addressed using mathematical models whose goals are twofold: 1) the assessment of global growth theories through hypotheses testing and 2) the forecast of future/past growth based on restricted number of data [2]. While the former is useful to uncover possible general mechanisms governing macroscopic tumor growth, the latter can be applied in preclinical drug development [3–5] or even for clinical personalized prediction of further disease development (examples of such applications can be found in [6,7] for prognosis of lung metastases, in [4,8,9] for studies on gliomas or in [10] for prostate cancer response to androgen suppression therapy). Cancer modeling offers a wide range of mathematical models (see [11] for a historical perspective) that can be classified according to their scale, approach (bottom-up versus top-down) or integration of spatial structure. At the cellular scale, agent-based models (see for instance [12]) are well-suited for studies on tumor spheroids, while the tissue scale is better described by continuous partial differential equations like reaction-diffusion models [9,13] or continuum-mechanics based models [14,15]. Here we focus on scalar data of volume longitudinal development and will consider non-spatial models for macroscopic description of tumor kinetics, based on ordinary differential equations. A plethora of models exists at this scale (see [16]), which reduces tumor growth to its more essential components, the most essential being pure proliferation (leading to exponential growth). Observations of non-constant doubling time during tumor history [17] have led investigators to consider more elaborated models such as the widely used Gompertz model [18,19], logistic or generalized logistic models, power growth [20] or two phase models (exponential followed by linear, see [5]). More recently, new models integrated tumor

neo-angiogenesis in the modeling by considering a dynamical carrying capacity [3,21].

Existence of this broad class of models raises the question of their relative relevance for description and forecast of tumor growth within a given preclinical mouse model. Here we present a comparative study of these models for both their descriptive and predictive power.

Materials and Methods

Ethics statement

This study was performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. The protocol was approved by the Institutional Animal Care and Use Committee (Protocol: #P11-324). The institution is AAALAC accredited and every effort was made to minimize suffering to the mice involved.

Mice experiments

Cell culture

Murine Lewis lung carcinoma (LLC) cells, originally derived from a spontaneous tumor in a C57BL/6 mouse [22], were obtained from American Type Culture Collection (Manassas, VA). The LLC cells were cultured under standard conditions [22] in high glucose DMEM (Gibco Invitrogen Cell Culture, Carlsbad, CA) with 10% FBS (Gibco Invitrogen Cell Culture) and 5% CO_2.

Tumor Injections

C57BL/6 male mice were used with an average lifespan of 878 days [23]. At time of injection mice were 6 – 8 weeks old (Jackson Laboratory, Bar Harbor, Maine). Subcutaneous injections of 10^6 LLC cells in 0.2 ml phosphate-buffered saline (PBS) were performed on the caudal half of the back in anesthetized mice. Tumor size was measured regularly with calipers to a maximum of 1.5 cm^3 when mice were sacrificed and the tissues processed.

Tissue processing

Mice were sacrificed with a 0.6 ml intraperitoneal injection of 2,2,2-tribromoethanol at 20 mg/ml. Tissues to be frozen-sectioned were dissected and slow-frozen in OCT (Tissue Tek, Fisher Scientific, Pittsburgh, PA) in the gas phase of liquid nitrogen. Tissues to be paraffin-sectioned were placed in 10% formalin, processed by standard protocol (R. D. Lillie, Histopathologic technic and practical histochemistry. Blakiston Division, New York, 1965.), placed in cassettes, and paraffin-embedded. Paraffin-embedded tissues were cut into 4 μm slices, placed on positively charged slides (Fisher Scientific), and stained for hematoxylin and eosin (H&E) stain using standard

protocols (R. D. Lillie, Histopathologic technic and practical histochemistry. Blakiston Division, New York, 1965.).

Models

The simplest model used for description of tumor growth is exponential growth, representing proliferation of a constant fraction of the total number of cells with constant cell cycle length. Initial exponential phase can be assumed to be followed by a linear growth phase, as done in [5]. The associated differential equation for the volume rate of change (growth rate) is

$$\begin{cases} \frac{dV}{dt} = a_0 V, & t \leq \tau \\ \frac{dV}{dt} = a_1, & t > \tau \\ V(t=0) = V_0 \end{cases}$$

where coefficient a_0 is related to proliferation (it is the fraction of proliferative cells times the inverse of the cell cycle length), a_1 drives the linear phase, time τ is computed such that there is continuity of the derivative $\left(\tau = \frac{1}{a_0}\log\left(\frac{a_1}{a_0 V_0}\right)\right)$ and V_0 is the initial volume. We considered three models deriving from this formula: a) fixed initial volume $V_0 = 10^6\ cells\ =\ 1\ mm^3$ (number of injected cells) and no linear phase $(a_1 = +\infty)$, referred to hereafter as *exponential 1*, b) free initial volume and no linear phase, referred to as *exponential* V_0 and c) the full model, referred to as the *exponential-linear model*.

Other growth models were considered, with the intent to fit the data without letting the initial volume as a degree of freedom and fixing its value to the number of injected cells). Although the number of cells that actually take to form a tumor is probably lower than the number of injected cells (around 60-80%), we consider $1\ mm^3$ as a reasonable approximation considering that the order of magnitude of the total growth curve is $1000\ mm^3$. The *Gompertz model* is described by the following differential equation and initial condition

$$\begin{cases} \frac{dV}{dt} = aV\log\left(\frac{K}{V}\right) \\ V(t=0) = 1\ mm^3 \end{cases}$$

where a is a coefficient related to proliferation kinetics and K is the so-called carrying capacity (maximal reachable volume). This model is built-up to exhibit exponential decrease of the relative growth rate $\left(\text{defined as } \frac{1}{V}\frac{dV}{dt}\right)$. The Gompertz model

corresponds to the limit when α goes to zero of the *generalized logistic model* (providing parameter a is rescaled to $\frac{a}{\alpha}$):

$$\begin{cases} \frac{dV}{dt} = aV\left(1 - \left(\frac{V}{K}\right)^{\alpha}\right) \\ V(t=0) = 1\,mm^3 \end{cases}$$

When $\alpha = 1$, this model will be simply referred to as the *logistic model.* In this case the instantaneous probability of proliferation of a cell is proportional to $1 - \frac{V}{K}$, which corresponds to mutual competition between the cells, due to space or nutrients limitation, for instance.

One complexity step above is a model assuming a dynamic (time-dependent) carrying capacity and introduced in [3,21]. The underlying theory consists in interactions of the tumor with its vasculature, represented by the carrying capacity (CC) K. Additionally considering that stimulation of the carrying capacity is proportional to the tumor surface, this model writes

$$\begin{cases} \frac{dV}{dt} = aV\log\left(\frac{K}{V}\right) \\ \frac{dK}{dt} = bV^{2/3} \\ V(t=0) = 1\,mm^3, K(t=0) = K_0 \end{cases}$$

and will be referred to as the *dynamic CC model*.

Eventually we considered the following model, first introduced for tumor growth description in [20] and that will be referred to as the *power growth model*

$$\begin{cases} \frac{dV}{dt} = aV^{\gamma} \\ V(t=0) = 1\,mm^3 \end{cases}$$

The underlying theory is that only a subset of the cancer cells are cycling, that part being proportional to a power of the volume, and thus having a smaller, possibly fractional, dimension than the tumor itself, proposing the concept of subdimensionality of the proliferative tissue. For instance $\gamma = \frac{2}{3}$ could represent a proliferative rim limited to the surface of the tumor, while $\gamma = 1$ describes proliferative cells uniformly distributed within the neoplasm and recovers exponential growth. It should be noted that in the former case simple calculations show that the tumor radius (proportional to $V^{2/3}$) grows linearly in time. A more complex equation than the

power growth model that consists in adding a loss term proportional to the volume in the differential equation, known as the von Bertalanffy model [24], will not be considered here because it resulted, in our analysis, in higher Akaike Information Criterion than the mere power growth equation (1073 versus 1070).

Fit procedure and goodness of fit criteria

Weighted least-squares

Based on statistical analysis of the measurement error (see the Results section), we fitted the models to the data using weighted nonlinear least square minimization. The following objective was used

$$J(\beta) = \sum_{i=1}^{n} \frac{\left(Y_i - f(t_i, \beta)\right)^2}{\Sigma_i^2} \quad (1)$$

where $f(t_i, \beta)$ stands for the output of model f at time t_i with parameter set β, n is the number of measurements and Σ_i is the variance associated to measure error for Y_i given by $\Sigma_i = \begin{cases} \sigma {Y_i}^{\alpha}, & Y_i \geq V_m \\ \sigma V_m^{\alpha}, & Y_i < V_m \end{cases}$ (see Results). Minimization of the sum of squared residuals was performed using the built-in Matlab [25] function *lsqcurvefit* that is based on a trust-region reflective optimization algorithm. Bounds on parameters were imposed, but only to ensure positivity of the parameters. Values of the parameters used for initialization of the minimization algorithm are reported in the supplementary Table 1.

Normalized error (NE)

From the obtained fit, various indicators of the goodness of the fit can be defined. Based on our *a priori* analysis of the measurement error we define the normalized error as

$$NE_i = \left| \frac{Y_i - f(t_i, \beta)}{\Sigma_i} \right|$$

for measurement i. We then considered the median of these normalized errors over all the fitted data points, pooling all the animals together. The resulting criterion (median normalized, denoted mNE) quantifies, in number of standard deviations, the median relative position of the model simulation within the errorbars and is considered as a good indicator of the goodness of fit. Notice that this criterion is relative to our underlying error model.

Coefficient of determination

The coefficient of determination is defined by

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \left(Y_i - f(t_i)\right)^2}{\sum_{i=1}^{n} \left(Y_i - \overline{Y}\right)^2}$$

where $\overline{Y}$ is the time average of the data points. It quantifies how much of the variability in the data is described by the model and how better is the model than fitting the data by only a constant line equal to the average value.

Root mean squared error (RMSE)

Another classical goodness of fit criterion that also penalizes lack of parsimony of the model (i.e. too many parameters) is given by the Root Mean Squared Error (RMSE) defined by

$$RMSE = \sqrt{\frac{1}{n-p} J(\beta^*)}$$

where p is the number of parameters and β^* is the parameter set giving the best fit.

p-value

We performed statistical goodness of fit χ^2 test (performed using 20 bins) to evaluate if the weighted residuals from a given fit with a given model were normally distributed with a standard deviation being the one observed in the measurement error analysis ($\sigma = 0.21$).

Population approach, mixed-effect statistical models and a priori information

Mixed-effect statistical models

The approach we explained above for the fitting procedure is based on minimization of criterion defined by formula (1) for one given individual and does not account for the fact that all the individuals are part of a same population and thus should somehow react the same, although having a (possibly wide) inter-animal variability. The mixed-effect approach, implemented in the Monolix software [26], consists in pooling all the individuals together and estimate a global distribution of the model parameters in the population. It is based on maximizing the likelihood of a parameterized distribution, which is done using the SAEM algorithm (stochastic algorithm for global optimization). From the estimation of the global log-likelihood (LLH), standard goodness of fit scores are derived such as the Akaike Information Criterion (AIC), defined as

$$AIC = -2 * LLH + 2 * p$$

where p is the number of parameters in the model. Such criterion allows ranking of the models for their fitting power, taking into account parsimony. Overall, the results

we obtained using Monolix where similar to the ones we had using individual fits (see Table 2).

A priori information

For tumor growth forecast, we used a statistical population approach considering integration of *a priori* information on the population distribution of parameters. We integrated 10 additional mice from a different but similar study of subcutaneous LLC growth in C57/BL6 mice, pooled all the animals together and randomly divided them into two groups. For each model, on the first group, individual fits where performed using all the available data. This allowed us to derive mean and standard deviation of the models parameters within the population. We then used this information to penalize the sum of squared residuals used for fits of the second group individuals, in the following way

$$J_P(\beta) = \frac{1}{n}\sum_{i=1}^{n}\frac{\left(Y_i - f(t_i,\beta)\right)^2}{\Sigma_i^2} + \frac{1}{p}\sum_{j=1}^{p}\frac{\left(\beta_j - \widehat{\beta_j}\right)^2}{\omega_j^2}$$

where $\widehat{\beta_j}$ is the mean value of parameter β_j's population distribution, ω_j is its standard deviation and p is the number of parameters. This procedure was repeated 100 times (i.e. 100 random assignments of the total population between 10 "learning" animals and 10 "forecast" animals), this number being considered as sufficient to be in the convergence limit of the large numbers law (no significant difference between 20 and 100 replicates, $p > 0.2$ by Student's t-test).

Results

Measurement error

We estimated the measurement error due to imprecision when using calipers in assessment of the tumor volume. Some of the measurements, one per time point per cage (over a total of 12 time points and 8 cages), were done twice within a few minutes interval in order to estimate the variability in the measurement. This gave in total 133 measures for which we have information about the error, which can be analyzed by considering the following statistical representation

$$Y = Y_T + \Sigma\varepsilon$$

where Y stands for the measure of a tumor whose real volume is denoted by Y_T, ε is a reduced centered Gaussian random variable and Σ is the error variance. The two independent measures we performed, termed Y_1 and Y_2, are strongly correlated (Figure 1A, correlation coefficient $r = 0.98$), but statistical analysis rejects variance independent of volume ($p = 0.004$, χ^2 test) and simple proportional error is only weakly significant ($p = 0.083$, Figure 1B). Indeed, errors made on small tumors are underestimated when considered proportional to the volume due to the difficulty to

detect the edges of subcutaneous implants. On the other hand large measurement errors are overestimated with a proportional variance. To overcome these two issues we propose the following expression of the error variance

$$\Sigma = \begin{cases} \sigma Y^{\alpha}, & Y \geq V_m \\ \sigma V_m^{\alpha}, & Y < V_m \end{cases} \quad (2)$$

In this model variance of the error is proportional to a power Y^{α} of the volume for tumor volumes larger than V_m while volumes smaller than V_m have same error as measuring V_m. The proportionality coefficient is denoted by σ,. We explored various values of V_m and α and found $\alpha = 0.84, V_m = 83\ mm^3$ to be able to describe dispersion of the error ($p = 0.196$, Figure 1C). This yielded a value of $\sigma = 0.21$.

This result was further confirmed by a fitting analysis performed with Monolix software [26], in which the Akaike information criterion (AIC) criterion was found to be lower when using an error variance model proportional to a power of the volume, as compared to constant or proportional error models (see Methods for details about mixed-effect models and population approach that is implemented in Monolix).

These results allow precise quantification of the measurement error inherent to our data, which is a fundamental step towards the assessment of a model's descriptive power since it allows us to quantitatively determine whether the data we have could have been generated by the model.

Robustness and numerical identifiability of the models

Specific assumption

For the dynamic CC model, allowing the stimulation power to vary did not significantly improve fit performances and resulted in higher Akaike Information Criterion, justifying the assumption of a fixed power ($\gamma = \frac{2}{3}$).

Numerical identifiability

Numerical identifiability of the fits to the initial parameter set given to the trust-region reflective minimization algorithm used by Matlab was assessed by systematically varying initialization of the algorithm. Compact subset of the parameter space of length 4 standard deviations above and below a baseline mean value (obtained by an *a priori* fit) in each parameter direction was meshed (11 discretization steps for each direction) and explored, with a total of $10{\times}11^P$ individual fits performed for each model, with p being the number of parameters in the model. We report in Table 1 results of sensitivity scores, defined as the fraction of fits that converged to the same parameter set as the baseline value (within a range of 10% error). When global numerical identifiability was not observed, further study was performed and the resulting variation of the best-fit parameter sets computed (Table 1).

The dynamic CC model exhibited a low identifiability score of 61.5% (out of 13310 fits). This is mostly due to its bi-dimensional nature (volume and carrying capacity are variables) with only one observable used for the fits. This fact results in variability mostly in estimation of K_0 with large deviations although significant variations were obtained also in the other parameters. The generalized logistic model had very low identifiability score (1.19%), mostly due to high volatility of parameter α (median deviation 390%) that confirms tendency of this parameter to be close to zero in our data and suggests the Gompertz as more adapted in the class of sigmoid-like models.

All the other models exhibited very good robustness in the parameter estimation, specifically the power growth model for which 100% of the fits gave the same optimal parameter set as the one derived with the baseline initial guesses, suggesting global numerical identifiability of these models.

Goodness of fit

We tested all the models for descriptive properties and quantified the goodness of fit on the total population of growth kinetics according to various criteria (see supplementary material for their definition). Results are reported in Table 2 and Figure 2.

The exponential 1 model was not able to describe the kinetics of our data (Figure 2B, Table 2) while allowing the initial volume to be a free parameter resulted in much better descriptive properties (Figure 2B). However, this resulted in a significantly higher initial volume V_0 than the number of injected cells ($23.4 \pm 7.47\ mm^3$ in the fits versus $1\ mm^3$ injected) and thus provided a biologically unrealistic description of tumor growth. Hence, in the following we did not allow for a free V_0 arguing that: a) it results in implausible initial volumes and b) models are usually very sensitive to this initial condition as small difference propagates exponentially in time. Consequently, for all the other models we fixed $V_0 = 1\ mm^3$.

Although the exponential-linear model exhibited good fits to the data in [5] it was not able to accurately describe ours (Figure 2B) as we don't observe stabilization of the growth in a linear regimen. Similarly, the logistic model had poor descriptive power (Figure 2B and Table 2). The three models exponential V_0, exponential-linear and logistic can be grouped together regarding to their fit properties, as models with approximate descriptive power. Indeed, these three models all failed the statistical χ^2 goodness of fit test ($p < 0.001,\ \text{Table 2}$). It should also be noted that the R^2 score appears as a coarse value of the fitting power because, although these models have R^2 values larger than 0.95, fits remain unsatisfactory as expressed by a median value of the normalized error larger than 1.

All the goodness of fit criteria we considered globally performed similarly and allowed us to rank the models (in ascending order for AIC in Table 2) for their descriptive power. It is worth noting that goodness of fit results were globally consistent, whether

using the Matlab built-in algorithm for minimization of least squares (based on a deterministic descent method) or the stochastic approximation of expectation maximization (SAEM) algorithm for likelihood maximization implemented in Monolix and based on a mixed-effect statistical population approach.

Four models (dynamic CC, Gompertz, power growth and generalized logistic) did not fail the test of Gaussian distribution of residuals with standard deviation given by the measurement error analysis (p>0.05, Table 2) and fell within the error bar for more than half of all the data points ($mNE < 1$, Table 2). Apart from the generalized logistic growth that had relatively elevated mean root mean square error (higher than exponential V_0), the other three models exhibited excellent scores for all the criteria considered; the slight differences between them did not allow discrimination between them. Parameter α from the generalized logistic was globally estimated to a very low value (0.06 ± 0.13, Table 3), suggesting a trend toward the Gompertz model since the generalized logistic converges to the Gompertz when α goes to 0. Indeed, the Gompertz has better AIC and mean RMSE as it performed better with fewer degrees of freedom.

Interestingly, two of the “high descriptive power” models have only 2 parameters (power growth and Gompertz); suggesting that the data we dispose intrinsically has less than two degrees of freedom.

Despite the complexity of internal cell populations and tissue organization (Figure 2A), these results show that at the macroscopic scale tumor growth exhibits relatively simple dynamic that can be captured through mathematical models. However, not all of them are equal in terms of descriptive power. Exponential based and logistic models have to be rejected while the power growth, Gompertz, dynamic CC and generalized logistic could reasonably have generated our data. These last four models are indistinguishable from one another and represent valid mathematical theories for the description of *in vivo* tumor growth.

Insights on macroscopic growth laws from mathematical models

From the observation that fitting to the exponential function requires assuming an unrealistically large initial volume parameter V_0, we deduce that growth occurs at a faster rate at initiation than at later time points. This is substantiated by the observation that computed growth rates from the exponential V_0 model are significantly lower than reported for the *in vitro* growth rate of LLC cells ($a = 0.233 \pm 0.0164\ day^{-1}$ in our analysis versus $1.12 \pm 0.12\ day^{-1}$ *in vitro* in [27]). Furthermore, although exponential growth gives a reasonable first-order approximation of established growth, superior fitting power of models exhibiting a decrease of the relative growth rate such as the Gompertz, generalized logistic or dynamic CC (Table 2) suggests non-constant doubling time over growth history. Although the logistic model does incorporate a non-constant doubling time feature, it did not accord with the data, while the Gompertz model did. This suggests that the underlying hypothesis

of the logistic model - a slowdown of the relative growth rate due only to competition between the cancer cells – is not sufficient to explain the observed growth. On the other hand, gompertzian growth (i.e., exponential decay of the relative growth rate) appears suitable as a descriptive model for this slowdown, although not providing any biological insight into the mechanistic basis for this observation. All the four best descriptive models, power growth, Gompertz, dynamic CC and generalized logistic, exhibited very similar relative growth rate decay profiles (supplementary Figure 1), emphasizing a possible general law in the way the fraction of proliferative cells decreases within tumors over time.

Due in parts to excellent identifiability of the model, the power growth parameters resulting from the fit exhibit low inter-animal variability, in particular coefficient γ.

The power growth model offers a theory that integrates linear growth of the tumor diameter, reported for instance in the case of gliomas [28]. This situation naturally occurs when $\gamma = \frac{2}{3}$ and describes proliferative cells limited to the surface of the lesion. On the other hand, exponential growth is recovered when $\gamma = 1$. However the model does not limit to these two extreme cases as any fractional power can occur. In our results, the fractal dimension of proliferative tissue was found to lie between 2 and 3 ($\gamma = 0.74 \pm 0.05$, giving a fractal dimension of 2.23 ± 0.15). This observation is substantiated by experimental results [3] where proliferative cells were found within the hypoxic regions of experimental tumors lying in the interior of the tumors. Our value of γ also matches remarkably well the values reported in [20] where the power growth model was shown to accurately fit to growth data of 300 C3H mouse mammary tumors (95% confidence interval of 0.73 ± 0.08 in their analysis versus 0.75 ± 0.03 in ours), suggesting a similar fractal pattern for the two different cell lines. Moreover, the relative growth rate at injection (given by the other model parameter, a) that we inferred from the fits is in remarkable agreement with its *in vitro* counterpart, the proliferation rate of LLC cells ($1.11 \pm 0.23\ day^{-1}$ in our analysis versus $1.12 \pm 0.12\ day^{-1}$ *in vitro* [27], a value confirmed by proliferation kinetics observed in our laboratory). Hence the power growth model offers not only good description of the growth curves of subcutaneous tumor growth but also a biologically meaningful law for neoplastic development with relevant coefficient values. Based on a parsimonious expression, it reproduced the observed and already reported decay of the relative growth rate and elucidates the phenomenon in terms of a plausible biological explanation.

The power growth model thus offers a valid and simple theory of macroscopic neoplastic growth that highlights a possible fractal nature of the proliferative tissue.

Forecasting tumor growth

The three highly descriptive models power growth, Gompertz and dynamic CC, were further assessed for their predictive power. Adjunction of the exponential V_0 model

was also considered. The challenge considered was to estimate future growth based on a given number of data points.

Models' predictive power

Despite similar descriptive properties (Table 2), the models did not perform equally well with regard to prediction (Figure 3). The first setting we considered was to predict future growth based parameters fitted using six data points. Goodness of the prediction was quantified by the median normalized error between model predictions and data, over all the future data points (Table 3). We also tested the predictability of the next data point (Table 3). Figure 3 illustrates the predictive performances of all the models for a given mouse. Model prediction was considered successful when median NE over the remaining data points was lower than three, corresponding to a model prediction within three standard deviations of the measurement error of the data. This methodology was considered in agreement with direct visual assessment of the goodness of prediction (see supplementary Figure 2 where all the individual dynamic CC predictions are presented). In this setting, the dynamic CC model gave good results, being able to predict 70% of the animals' future trends and 90% of them at horizon (depth) 1 (Table 4). Situations where the model failed to accurately predict occurred when unexpected growth acceleration (animal 4 in supplementary Figure 2) or deceleration (animal 6 in supplementary Figure 2) occurred. On the other hand, the good descriptive properties of the Gompertz model did not translate into accurate predictive power since less than 50% of the individuals could be globally predicted (Table 4). The power growth model exhibited satisfactory predictive power (60% success rate for global future growth and 70% at horizon 1, see Table 4) despite its low number of parameters.

Prediction depth

For evaluation of the global predictive properties of the models, predictions using a variable number of given data points and varying prediction depth were performed and are shown in Figure 4 and Table 4. The number of data points used in a particular setting is denoted by N, with N ranging from 3 to 7. We then asked the models to predict the d-th next data point (i.e. the $N+d$-th one), with the prediction depth d ranging from 1 to 7. The associated success rate for a given model was denoted S_d^N. The results are summarized in Figure 4. Only two models had an overall median success rate higher or equal to 50%, namely dynamic CC (55%) and power growth (60%). The better median overall success rate of the power growth model is compensated by a lower mean overall success rate (48.3% versus 51.3%) indicating that, in those situations where the dynamic CC model was able to predict, it had a higher success rate. This happened for instance when $N=4$ or $N=6$. Globally, for the power growth and dynamic CC models, the high descriptive accuracy translated into substantial predictive power. Not surprisingly, the success rate increases when the models are fed with more data points or when the prediction depth is reduced.

In contrast, despite its very good descriptive properties (Table 2 and Figure 2), the Gompertz model exhibited low global predictive power, with a median overall success of 27.5% (Table 4 and Figure 4) and a lower score than the power growth in all situations but one ($N = 6$ and $d = 1$).

Interestingly, the basic exponential model with free initial volume performed well at a short time depth, even reaching a 90% success rate for prediction of the next data point using only 3 data points, while power growth and dynamic CC had the same score of only 60% in this situation. This can be explained by the mathematical fact that locally every dynamical system is exponential. Moreover the exponential model cannot exhibit slowdown of the relative growth rate. Thus, while models such as the power growth and the dynamic CC interpret an initial slow growth (or even stable initial growth) as a possible very fast decrease of the growth rate, the exponential does not and is able to predict an unexpected larger fourth time point. However, when looking at depth 2, the exponential model looses its superiority with the success rate S_2^3 drastically falling to 50%, while the power growth is stable at 60% and dynamic CC falls to 50%.

Taken together, these results suggest the power growth and the dynamic CC models as appropriate candidates for tumor growth forecast. Within the setting considered in this section, predictions can be made with an accuracy level of approximately 60%, up to a depth of four days (corresponding to approximately five months in humans) with sufficient number (≥ 6) data points, while only a two days depth can be expected when fewer data points are used.

A priori information

When few number of data points are available, for example with only three, individual predictions based on fits only was shown to be globally poor, especially over a large time frame (Figure 4). However, this situation is likely to be the clinically relevant since few clinical examinations are performed before the beginning of therapy. An interesting statistical method consists in integrating *a priori* information in the distribution of the parameters, learned from a given database, and to combine this information with the individual estimation from the available data on a given animal. To do so, we integrated 10 more mice to the study to gain statistical power and then randomly divided the population of 20 animals between two groups. One group was used to learn the parameters distribution and the other for forecast purposes. It should be pointed out that for a given individual, no information from that individual was used to estimate the *a priori* distributions: only information from an independent group. The full procedure was replicated 100 times to ensure statistical significance.

Predictions obtained using this technique were significantly improved when forecasting global future growth curves based on 3 data points (Figure 5). Not all the models equally benefited from the addition of *a priori* information in the fitting procedure (Figure 5). Low-parameterized models such as power growth, exponential V_0 and Gompertz, which also have the lowest parameter inter-individual variability

(Table 3), exhibited great benefit, while the dynamic CC model (which has three degrees of freedom) only had modest but still significant benefit (from 19.7% to 32.4%). Indeed, due to a widely spread distribution of parameters (especially K_0, see Table 3), the *a priori* distribution does not contain much information for this model, and does not add much information to the fits. On the other hand the power growth model, whose distribution in γ is particularly narrow (Table 3), has a much more informative *a priori* distribution that translated into the most drastic improvement of the predictive power (from 14.9% to 60.2%). The impact of the addition of the *a priori* information was less important when using more data points (results not shown).

These results demonstrate that addition of *a priori* information in the fit procedure greatly improves the forecast performances of the models, in particular when using fewer data points with low-parameterized models such as the power growth model.

Discussion

Rigorous study of the descriptive and predictive power for a class of mathematical models of tumor growth was performed. Based on a detailed quantification of the measurement error, five of the nine models initially considered exhibited significant descriptive power. Numerical identifiability was also considered as a criterion for model comparison.

Derivation of a specific measure error model was a fundamental consideration in the quantitative assessment of the models' performances and statistical rejection of inaccurate growth theories. As already observed by others [17–19], our results confirmed that tumor growth cannot be continuously exponential (constant doubling time) and consequently that it cannot be explained only by proliferation of a constant fraction of the neoplasm, specifically at initiation. This fact has important implications in terms of identification of the inception time of the tumor [17].

Sigmoidal models such as the Gompertz model (exponential decay of the relative growth rate) were found to be valid descriptions of tumor growth, while logistic decay of the relative growth rate had to be rejected. The exponential-linear model was also rejected by our analysis, a result that contrasts with its good fitting power observed in [5] and that can be explained by the error model we used in our analysis (volume-dependent variance) that derived from our quantification of the measurement error. Using a constant variance error model gave a much a better rank to this model. Despite its good fitting properties, the Gompertz model does not elucidate the underlying biology and the dynamic CC model, while being more biologically-based, and found to have similar descriptive properties, exhibited large variability of the parameters in the population that translated into low numerical identifiability and ultimately limited predictive power. Nevertheless, the latter model was not designed with the intent to quantify tumor growth, but rather to be able to describe the effects of anti-angiogenic agents on global tumor dynamics.

As a result of our analysis, the power growth model (relative growth rate proportional to a power of the volume) appears as a simple, robust, descriptive and predictive mathematical model for murine tumor growth kinetics that has clear and simple biological foundation. It suggests a general law of macroscopic *in vivo* tumor growth: only a subset of the tumor cells proliferates, the measure of this subset being proportional to a constant fractional power of the volume (its fractal dimension). This model showed a close match to our data and to the Gompertz curve (in the range of the observed volumes) and reconciles the Gompertz model with the biology by giving a mechanistic explanation of the growth rate decay that naturally happens when fractal dimension of the proliferative tissue is lower than 3. Of all the models we considered, the power growth model is the most parsimonious and identifiable, as well as the most biologically explicit, descriptive, and predictive. Moreover, it fits well with the general concept of fractal growth that is ubiquitous in biology of growth processes where self-similarity arises from dynamical auto-organization (examples being trees and pulmonary or vascular development).

Nevertheless, the origin of the fractal nature of the proliferative tissue remains to be elucidated. This should also be taken with caution when dealing with very small volumes (at the scale of several cells for instance) for which the assumption of fractal dimension falls, since the tumor tissue cannot be considered as a continuous medium anymore. Additionally, our results were obtained in a particular setting (syngeneic murine model) and, although consistent with other results involving mammary cell line [20], remain to be confirmed and substantiated by extension to broader experimental (in particular human) settings.

Potential use of mathematical models as forecasting tools was assessed. On top of its very good descriptive properties and identifiability robustness, the power growth model was found to be globally the most predictive over a wide range of situations regarding to number of data points used and prediction depths, with an overall median success score of 60%. In some situations, other models are more indicated, such as the dynamic CC model when using 6 data points and predicting the next data point (90% success).

Use of *a priori* information to facilitate parameter identification proved helpful, particularly when few data were available and models had a low number of parameters. In such a situation, it improved the success rate of the power growth model from 14.9% to 60.2%. This comes partly from the important homogeneity of our growth data that generated a narrow and very informative distribution of the power growth parameter γ, which in turn powerfully assisted fitting procedure. In more practical situations such as patient data, much more heterogeneity of the growth data can be expected.

Translating our results into a clinical setting raises the possibility of forecasting solid tumor growth using simple macroscopic models and proposes the power growth model in particular as a good initial candidate. Further information could and should

be extracted from (functional) imaging devices, feeding more complex mathematical models that could help design more accurate *in silico* prediction tools [7,8].

Our analysis shows that use of mathematical models could also be a valuable tool for helping preclinical anti-cancer research as it could lead to interesting applications for assessing drug efficacy (for example, by comparing the treated growth curve to the expected growth when no treatment is administered). Although integration of therapy remains to be added (and validated) to the power growth model, more classical models (such as exponential-linear [5] or dynamic CC [3]) have already been shown to be able to predict cytotoxic or anti-angiogenic effects of drugs on tumor growth. Our methods have allowed precise quantification of their respective descriptive and predictive powers, which, in combination with the models' intrinsic biological foundations, could be of value when deciding among such models which best captures the observed growth behaviors relevant preclinical settings.

Acknowledgments

We thank Etienne Baratchart for valuable suggestions and comments. This work was supported by the National Cancer Institute under Award Number U54CA149233 (to L. Hlatky). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institute or the National Institutes of Health.

References

1. Gilewski T, Norton L (2010) Cytokinetics. In: Hong WK, Bast RCJ, Halt WN, Kufe DW, Pollock RE, et al., editors. Holland-Frei Cancer Medicine. PMPH-USA.

2. Gammon K (2012) Forecasting cancer. Nature 491: S66–S67.

3. Ribba B, Watkin E, Tod M, Girard P, Grenier E, et al. (2011) A model of vascular tumour growth in mice combining longitudinal tumour size data with histological biomarkers. Eur J Cancer 47: 479–490. doi:10.1016/j.ejca.2010.10.003.

4. Ribba B, Kaloshi G, Peyre M, Ricard D, Calvez V, et al. (2012) A tumor growth inhibition model for low-grade glioma treated with chemotherapy or radiotherapy. Clin Cancer Res 18: 5071–5080. doi:10.1158/1078-0432.CCR-12-0084.

5. Simeoni M, Magni P, Cammia C, De Nicolao G, Croci V, et al. (2004) Predictive pharmacokinetic-pharmacodynamic modeling of tumor growth kinetics in xenograft models after administration of anticancer agents. Cancer Res 64: 1094–1101. doi:10.1158/0008-5472.CAN-03-2524.

6. Colin T, Iollo A, Lombardi D, Saut O (2010) Prediction of the Evolution of Thyroidal Lung Nodules Using a Mathematical Model. ERCIM News: 37–38.

7. Cornelis F, Saut O, Cumsille P, Lombardi D, Iollo A, et al. (2013) In vivo mathematical modeling of tumor growth from imaging data: Soon to come in the future? Diagn Interv Imaging 94: 593–600. doi:10.1016/j.diii.2013.03.001.

8. Baldock a L, Rockne RC, Boone a D, Neal ML, Hawkins-Daarud A, et al. (2013) From patient-specific mathematical neuro-oncology to precision medicine. Front Oncol 3: 62. doi:10.3389/fonc.2013.00062.

9. Wang CH, Rockhill JK, Mrugala M, Peacock DL, Lai A, et al. (2009) Prognostic significance of growth kinetics in newly diagnosed glioblastomas revealed by combining serial imaging with a novel biomathematical model. Cancer Res 69: 9133–9140. doi:10.1158/0008-5472.CAN-08-3863.

10. Portz T, Kuang Y, Nagy JD (2012) A clinical data validated mathematical model of prostate cancer growth under intermittent androgen suppression therapy. AIP Adv 2: 011002. doi:10.1063/1.3697848.

11. Araujo RP, McElwain DLS (2004) A history of the study of solid tumour growth: the contribution of mathematical modelling. Bull Math Biol 66: 1039–1091. doi:10.1016/j.bulm.2003.11.002.

12. Gao X, McDonald JT, Hlatky L, Enderling H (2013) Acute and fractionated irradiation differentially modulate glioma stem cell division kinetics. Cancer Res 73: 1481–1490. doi:10.1158/0008-5472.CAN-12-3429.

13. Gatenby R a, Gawlinski ET (1996) A reaction-diffusion model of cancer invasion. Cancer Res 56: 5745–5753.

14. Ambrosi D, Mollica F (2003) Mechanical Models in Tumour Growth. In: Preziosi L, editor. Cancer Modelling and Simulation. CRC Press. pp. 142–166.

15. Bresch D, Colin T, Grenier E, Ribba B, Saut O (2010) Computational Modeling of Solid Tumor Growth: The Avascular Stage. SIAM J Sci Comput 32: 2321. doi:10.1137/070708895.

16. Gerlee P (2013) The model muddle: in search of tumor growth laws. Cancer Res 73: 2407–2411. doi:10.1158/0008-5472.CAN-12-4355.

17. Steel GG, Lamerton LF (1966) The growth rate of human tumours. Br J Cancer 20: 74–86.

18. Laird AK (1964) Dynamics of tumor growth. Br J Cancer 13: 490–502.

19. Norton L (1988) A Gompertzian model of human breast cancer growth. Cancer Res 48: 7067–7071.

20. Dethlefsen L a, Prewitt JM, Mendelsohn ML (1968) Analysis of tumor growth curves. J Natl Cancer Inst 40: 389–405.

21. Wilson S, Grenier E, Wei M, Calvez V, You B, et al. (2013) Modeling the synergism between the anti-angiogenic drug sunitinib and irinotecan in xenografted mice. PAGE 22. p. 2826.

22. Bertram JS, Janik P (1980) Establishment of a cloned line of Lewis Lung Carcinoma cells adapted to cell culture. Cancer Lett 11: 63–73.

23. Kunstyr I, Leuenberger HG (1975) Gerontological data of C57BL/6J mice. I. Sex differences in survival curves. J Gerontol 30: 157–162.

24. Bertalanffy L von (1949) Problems of organic growth. Nature 163: 156–158.

25. The Mathworks (2012) Matlab.

26. Monolix software (2013).

27. Pyaskovskaya ON, Kolesnik DL, Kolobov A V, Vovyanko SI, Solyanik GI (2008) Analysis of growth kinetics and proliferative heterogeneity of Lewis lung carcinoma cells growing as unfed culture. Exp Oncol 30: 269–275.

28. Neal ML, Trister AD, Ahn S, Baldock A, Bridge C a, et al. (2013) Response classification based on a minimal model of glioblastoma growth is prognostic for clinical outcomes and distinguishes progression from pseudoprogression. Cancer Res 73: 2976–2986. doi:10.1158/0008-5472.CAN-12-3588.

Figure 1: **Volume measurement error**. A. First measured volume Y_1 against second one Y_2. Largest (L) and smallest (w) diameters were measured subcutaneously using calipers and then formula $V = \frac{\pi}{6} w^2 L$ was used for computation of the volume (ellipsoid). Also plotted is the regression line (correlation coefficient $r = 0.98$, slope of the regression $= 0.96$). B. Error $Y_1 - Y$ against approximation of the volume given by the average of the two measurement $Y = \frac{Y_1 + Y_2}{2}$. The χ^2 test rejected Gaussian distribution of constant variance ($p = 0.004$) C. Histogram of the normalized error applying the error variance model given by $\Sigma = \begin{cases} \sigma Y^{\alpha}, & Y \geq V_m \\ \sigma V_m^{\alpha}, & Y < V_m \end{cases}$, with $\alpha = 0.84$ and $V_m = 83\ mm^3$. It shows Gaussian distribution ($p = 0.196$) with standard deviation $\sigma = 0.21$.

Figure 2: Descriptive power. A. Excised tumor and hematoxylin and eosin immunostaining of the tissue resulting from *in vivo* growth of LLC cells. B. Illustrative example of all growth models fitting the same individual kinetic. Errorbars correspond to the standard deviation of the *a priori* estimate of measurement error. From visual examination on this example, exponential 1, Logistic and exponential-linear are not

appropriate while the others describe the growth in a satisfactory manner. C. Distributions of normalized errors (NE) for the tested models. Residuals include fits over all the animals and all the time points. Log = generalized, Exp-L = exponential-linear, GLog = generalized logistic, Dyn CC = dynamic CC, Gomp = Gompertz, PG = power growth.

Figure 3: Predictive power. Illustrative example of the forecast performances of the four best descriptive models. Six data points were used to learn the animal parameters and predict future growth. Although all models succeeded in predicting the next day data, exponential V_0 and Gompertz failed for global forecast of the future while power growth and dynamic CC succeeded, for this animal.

Figure 4: Prediction depth. Test of the predictive power of the models depending on the number of data points used and the prediction depth in the future. At position (N, d) the color represents percentage of successfully predicted animals when using N data points and forecasting the $N + d$ -th data. Only data where $N + d \leq 10$ was considered of interest since few animals had more than 10 longitudinal measurements. First line is model with two parameters and second line models with three parameters.

Figure 5: *A priori* information. Global group of 20 animals was randomly divided into two subgroups. One was used to retrieve the models parameters distribution in the population. This *a priori* information was then combined to the individual information from fitting $N = 3$ data points in order to predict all the future growth. The full procedure was repeated 100 times. For a given individual, prediction was considered successful when mean normalized error was lower than 3. A. Success rates of the models over all replicates when using *a priori* information on the parameters population distributions, $N = 3$ data points and predicting the global future growth curve (mean $\pm$ standard deviation). Improvement of the prediction is statistically significant in all situations ($p < 10^{-18}$ by Student's t-test) B. Illustrative example of the benefit of adjunction of *a priori* information for a given mouse, using the power growth model for prediction.

Supporting Information Legends

Supplementary Figure 1: **Relative growth rates.** Comparison of the time decrease of growth rates the considered models. Parameters are the ones resulting from the individual fit of tumor growth curve of an individual mouse. Note that the five best descriptive models (Gompertz, dynamic CC, power growth and generalized logistic) have very similar profiles while the others exhibit qualitatively different behavior.

Supplementary Figure 2: Individual predictions of future growth using 6 data points and the dynamic CC model. Individual parameters were estimated using the 6 first data points and future growth is extrapolated. Based on the criterion of a median

NE smaller than 3 for the total future prediction (meaning that the median model prediction is within 3 standard deviations of the measurement error), 7 tumor growths were considered to be reasonably predicted, corresponding to animals 1, 2, 5, 7, 8, 9 and 10. Prediction of only the next data point was considered successful for all but animal 4.

Supplementary Figure 3: Forecast improvement of the Power Growth model when using *a priori* information. Fits are performed using the first three data points, for each animal. *A priori* information (learned on a different data set) is added during the fit procedure for the predictions on the right. Shown is a particular replicate among the 100 subdivisions of the global group (20 mice) into one "learning" group and one "forecast" group.

Model	Identifiability score (%)	Par.	Score (%)	Median Dev. (%)
Power growth	100	-	-	-
Gompertz	99.8	a	100	-
		K	99.8	0.001
Dynamic CC	61.5	a	86.4	0.96
		b	94.0	0.4
		K_0	75.1	2.28
Generalized logistic	1.19	a	43.1	87.2
		K	67.6	72.5
		α	25.5	390
Exponential V_0	100	-	-	-
Exponential-linear	95.3	a_0	95.3	0.0013
		a_1	100	-
Logistic	100	-	-	-
Exponential 1	100	-	-	-

Table 1: Robustness of the models' numerical identifiability. Numerical identifiability of the models was assessed by systematically varying the initial condition of the optimization algorithm in a range of diameter 4 standard deviations in each parameter direction. We then tested agreement with the parameter sets obtained with base value of the initial guess. An identifiability score was then derived by computing the proportion of fits giving different convergence of the minimization, the difference being defined by a relative deviation larger than 10%. The identifiability score reported is the proportion of success, among the N^p individual fits where p is the number of parameters and N is the number of meshes in each parameter direction (here $N = 11$). When lower than 100%, further analysis was performed and the same score was computed for each parameter of the model. We also report their median relative deviation. Par. = Parameter. Dev. = Deviation

Model	**p**	**AIC**	**mNE**	**mR^2**	**mRMSE**	**J**	**Number of parameters**
Dynamic CC	0.721	1054	0.81	0.98	1.73	9.38	3
Gompertz	0.636	1069	0.74	0.98	1.83	12.6	2
Power growth	0.849	1070	0.87	0.98	1.72	10.8	2
Generalized logistic	0.427	1071	0.89	0.98	2.15	14.6	3
Exponential V_0	<0.001	1081	1.23	0.95	2.02	14.7	2
Logistic	<0.001	1098	1.63	0.96	2.90	28.9	2
Exponential-linear	<0.001	1154	1.46	0.96	2.80	27.6	2
Exponential 1	<0.001	1267	5.78	0.63	6.36	147	1

Table 2: **Fitting performances of growth models**. p = p-value of the χ^2 test for normal distribution of residuals with standard deviation $\sigma = 0.21$. AIC = Akaike Information Criterion computed using Monolix (proportional power variance error model). mNE = median normalized error (over all time points and animals). Normalized error is defined by $NE_i = \left|\frac{Y_i - f(t_i,\beta)}{\Sigma_i}\right|$ for model f, parameter set β, time point t_i, data Y_i and Σ_i defined by formula (2). mR^2 = coefficient of determination averaged over all the individual fits. mRMSE = Root Mean Squared Errors averaged over all the individual fits. J = total sum of squared errors upon all the individuals.

Model	Parameters (CV)		
Power growth	$\boldsymbol{a}$ $(\boldsymbol{day^{-1}mm^{\gamma-1}})$		$\boldsymbol{\gamma}$
	1.11 (29)		0.74 (6.5)
Gompertz	$\boldsymbol{a}$ $(\boldsymbol{day^{-1}})$		$\boldsymbol{K}$ $(\boldsymbol{mm^3})$
	8.87×10^{-2} (26)		8.92×10^{3} (67)
Dynamic CC	$\boldsymbol{a}$ $(\boldsymbol{day^{-1}})$	$\boldsymbol{b}$ $(\boldsymbol{mm^{\frac{3}{2}}\cdot day^{-1}})$	$\boldsymbol{K_0}$ $(\boldsymbol{mm^3})$
	9.33×10^{-1} (109)	2.91 (43)	44.8 (115)
Generalized logistic	$\boldsymbol{a}$ $(\boldsymbol{day^{-1}})$	$\boldsymbol{K}$ $(\boldsymbol{mm^3})$	$\boldsymbol{\alpha}$
	158 (78)	7.49×10^{3} (38)	0.055 (237)
Exponential V_0	$\boldsymbol{a}$ $(\boldsymbol{day^{-1}})$		$\boldsymbol{V_0}$ $(\boldsymbol{mm^3})$
	0.23 (7)		17.0 (29)

Table 3: **Parameter values for the four best descriptive models + exponential $\boldsymbol{V_0}$.** Shown is the mean value within the population and in parenthesis the coefficient of variation (CV, defined as the (standard deviation divided by mean and multiplied by 100) that quantifies inter-individual variability.

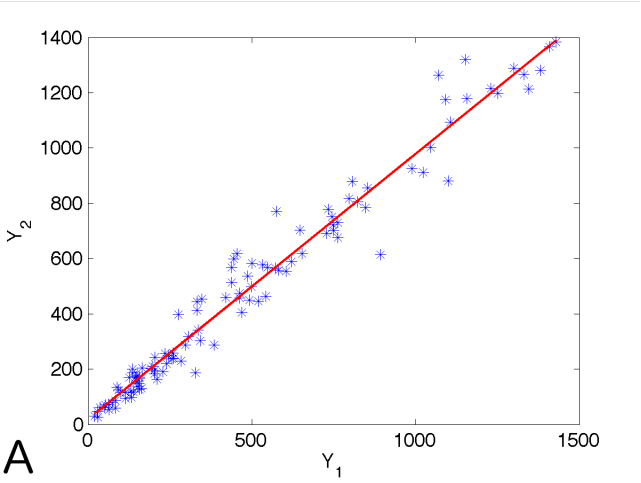
Model	Overall median success (%)	Overall mean success (%)	$\boldsymbol{S^6_{glob}}$ (%)	$\boldsymbol{mNE^6_{glob}}$	$\boldsymbol{S^6_1}$ (%)
Power growth	60	48.3	60	2.22	70
Dynamic CC	55	51.3	70	1.89	90
Exponential V_0	45	42.8	30	4.24	60
Gompertz	27.5	36.8	40	3.92	80

Table 4: **Forecast properties of four growth models having equivalent descriptive power plus exponential growth**. Prediction was considered successful when normalized error NE (or median NE when prediction involves more than one data point) was lower than 3.The first column is global quantification of the predictive power over all the number of data points used (from 3 to 7) and all the prediction depths (up to 7). The last three columns are quantifications of the predictive goodness for one given number of data points (N=6). Predictions were assessed on the remaining data points. S^6_{glob} = success rate for prediction of the global remaining curve, based on 6 data points (total of 10 mice). mNE^6_{glob} = median normalized prediction error among all the future growths of all individuals. S^6_1 = success rate for prediction only of the next data point, i.e. a prediction depth of one.

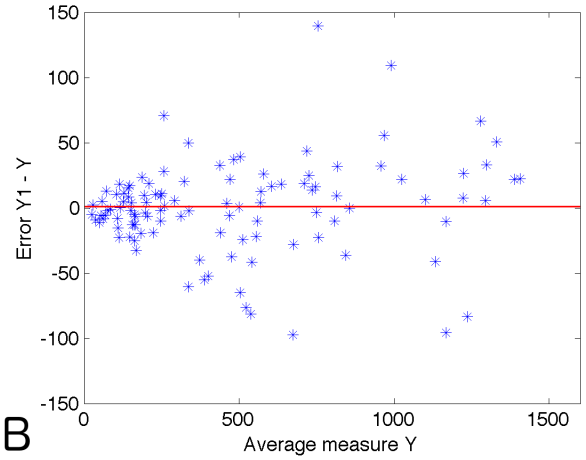
Model	Parameter	Initialization
Power growth	a	1
	γ	2/3
Gompertz	a	0.1
	K	10000
Dynamic CC	a	3
	b	0.5
	K_0	10
Generalized logistic	a	10
	K	10000
	α	0.01
Exponential V_0	a	0.1
	V_0	20
Exponential-linear	a_0	0.2
	a_1	500
Logistic	a	1
	K	10000
Exponential 1	a	0.1

Supplementary Table 1: Initializations of the least squares minimization algorithm

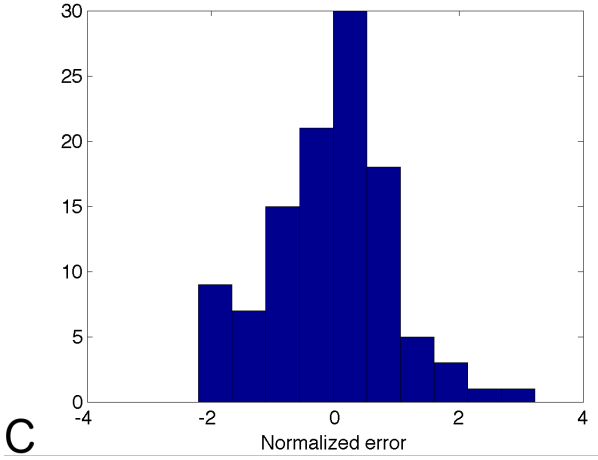
Figure 1


1400
1200
1000
800
600
400
200
0
0
500
1000
1500
Y2
Y1

A

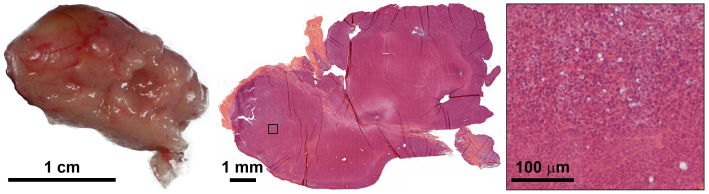

150
100
50
0
-50
-100
-150
0
500
1000
1500
Error Y1 - Y
Average measure Y

B

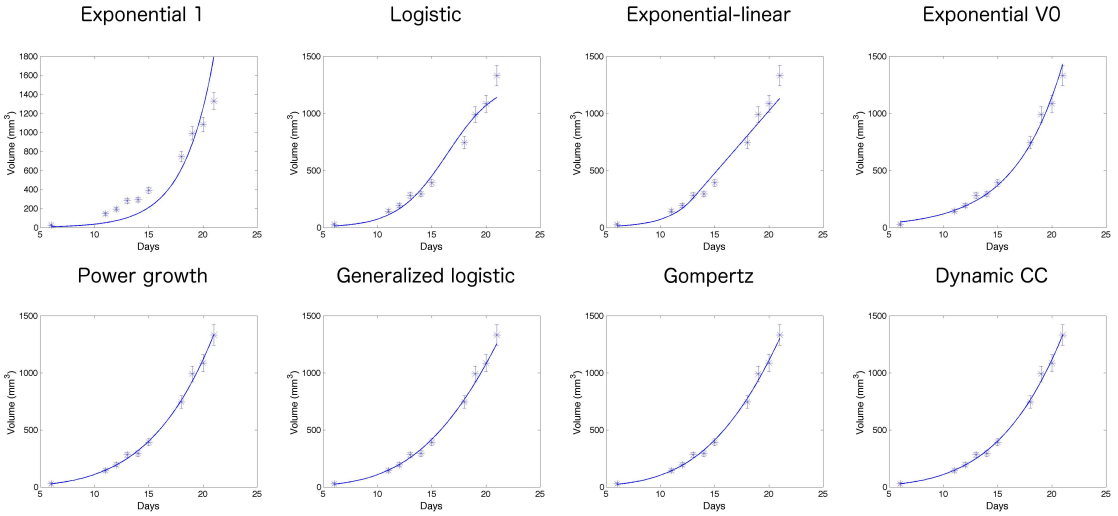

30
25
20
15
10
5
0
-4
-2
0
2
4
Normalized error

C

A


1 cm
1 mm
100 μm

B


Exponential 1
Logistic
Exponential-linear
Exponential V0
Power growth
Generalized logistic
Gompertz
Dynamic CC
Volume (mm3)
Days

C

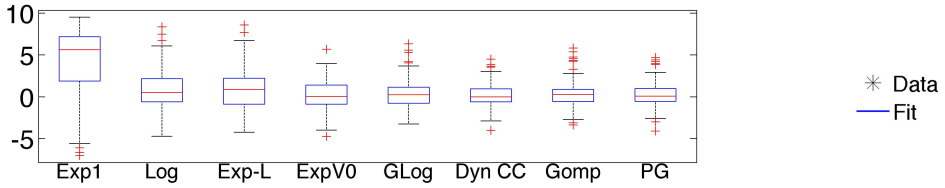

10
5
0
-5
Exp1
Log
Exp-L
ExpV0
GLog
Dyn CC
Gomp
PG
Data
Fit

Figure 2

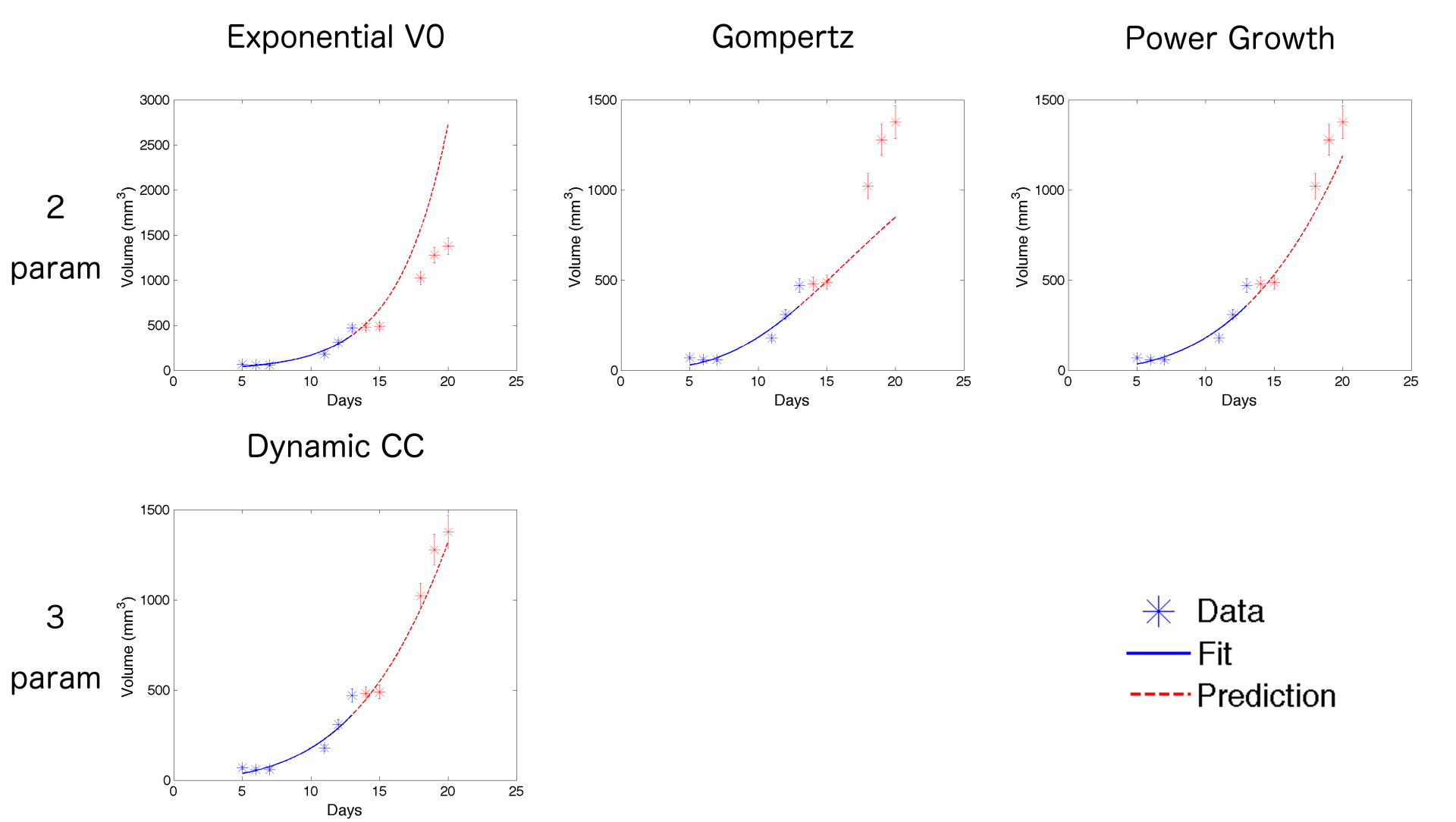

Exponential V0
Gompertz
Power Growth
2
param
Dynamic CC
3
param
Volume (mm3)
Days
Data
Fit
Prediction

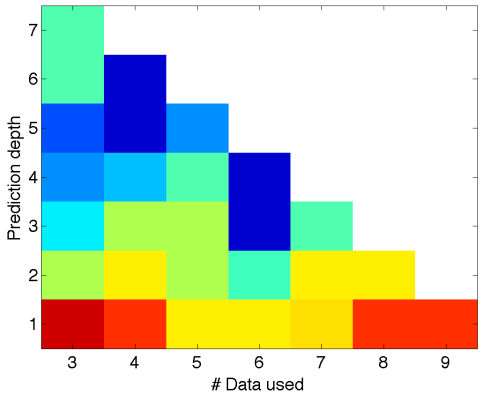
Figure 3

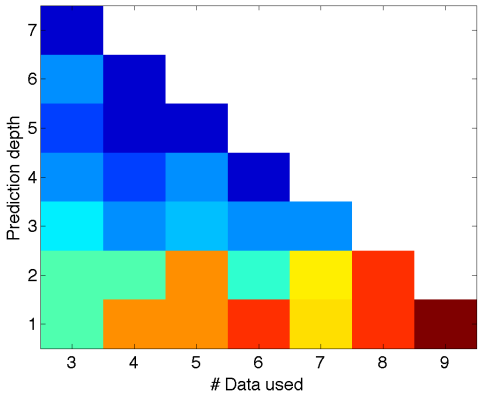
Exponential V0

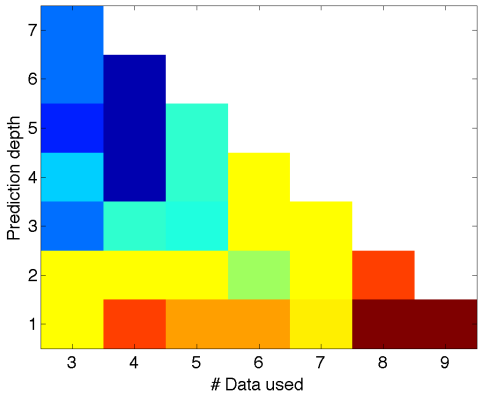
Gompertz

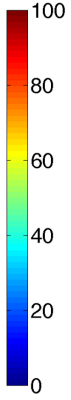
Power Growth

2 param









Dynamic CC

3 param

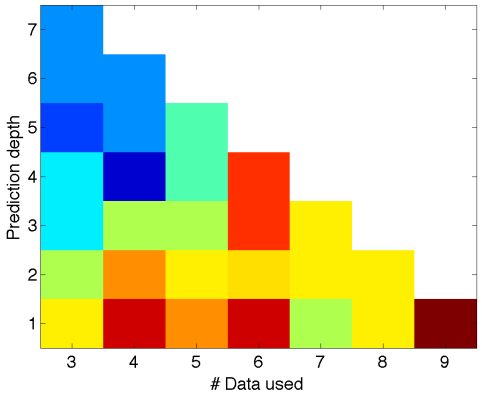


Figure 4

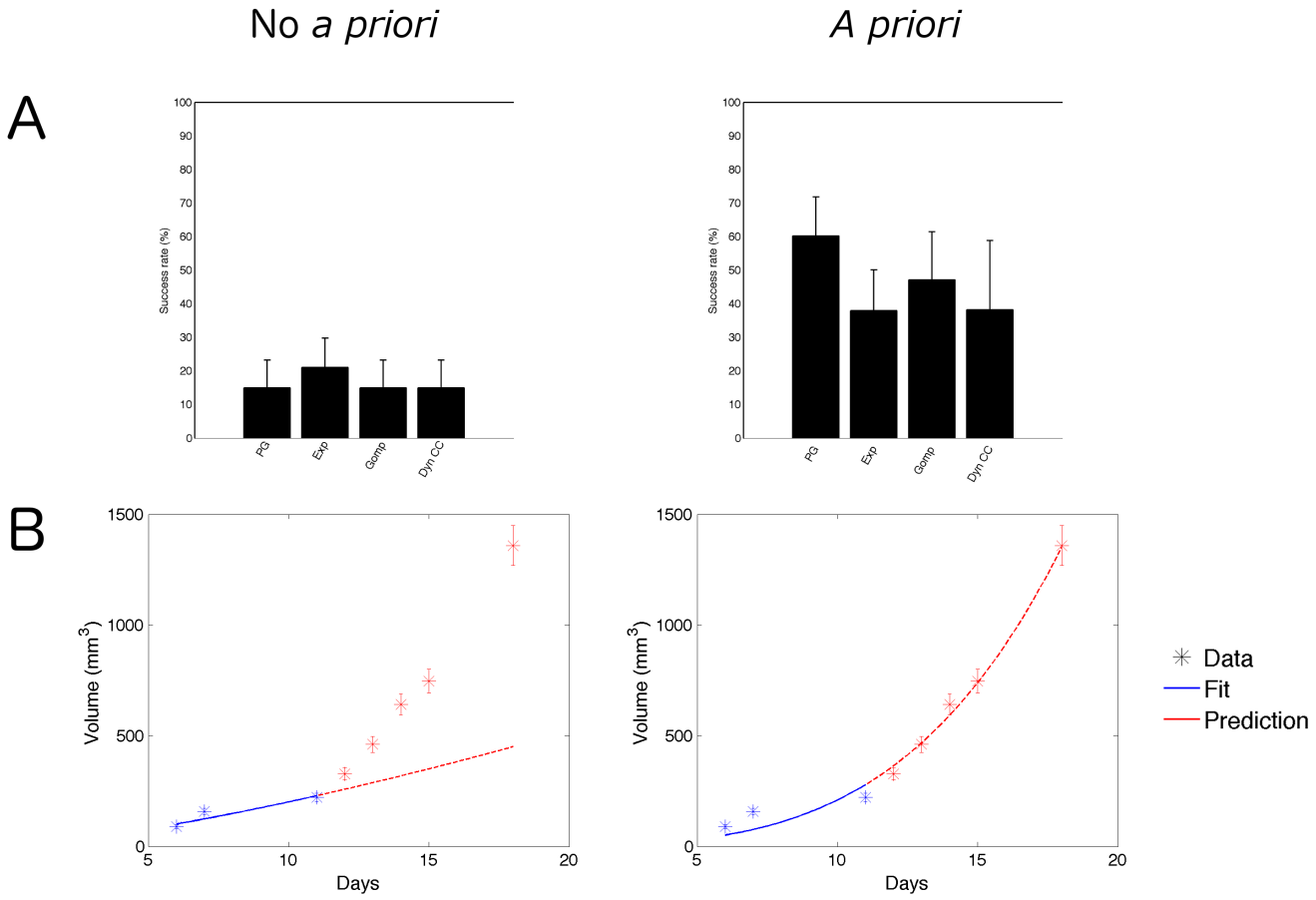
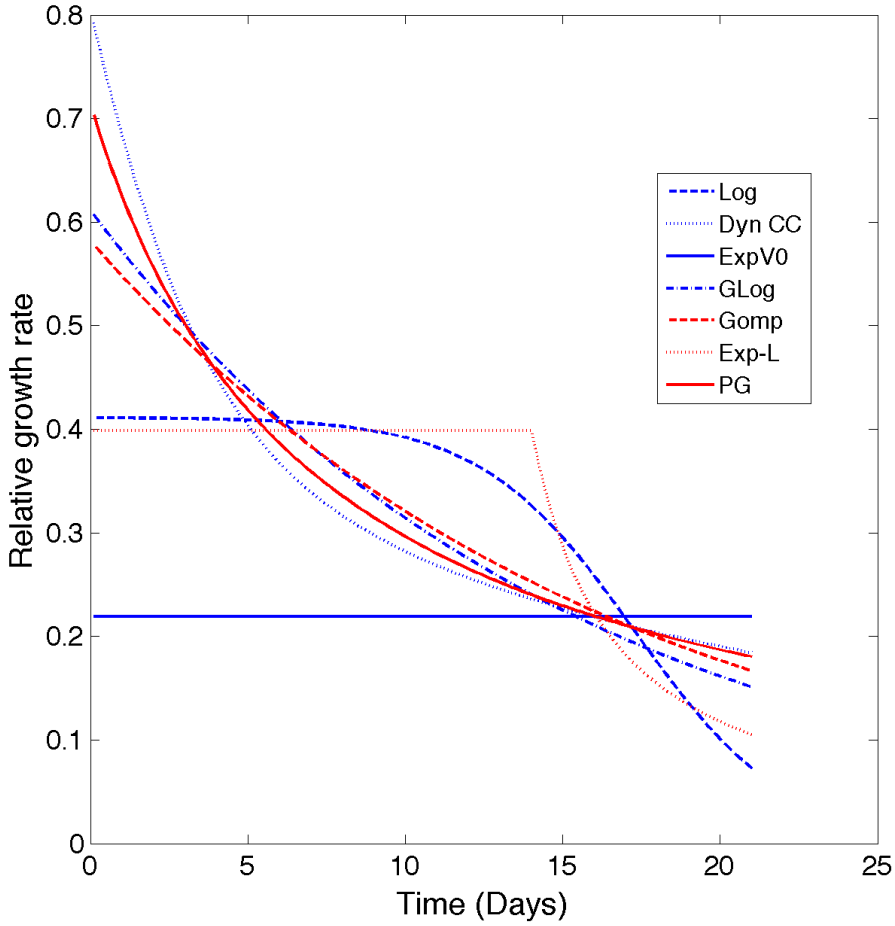
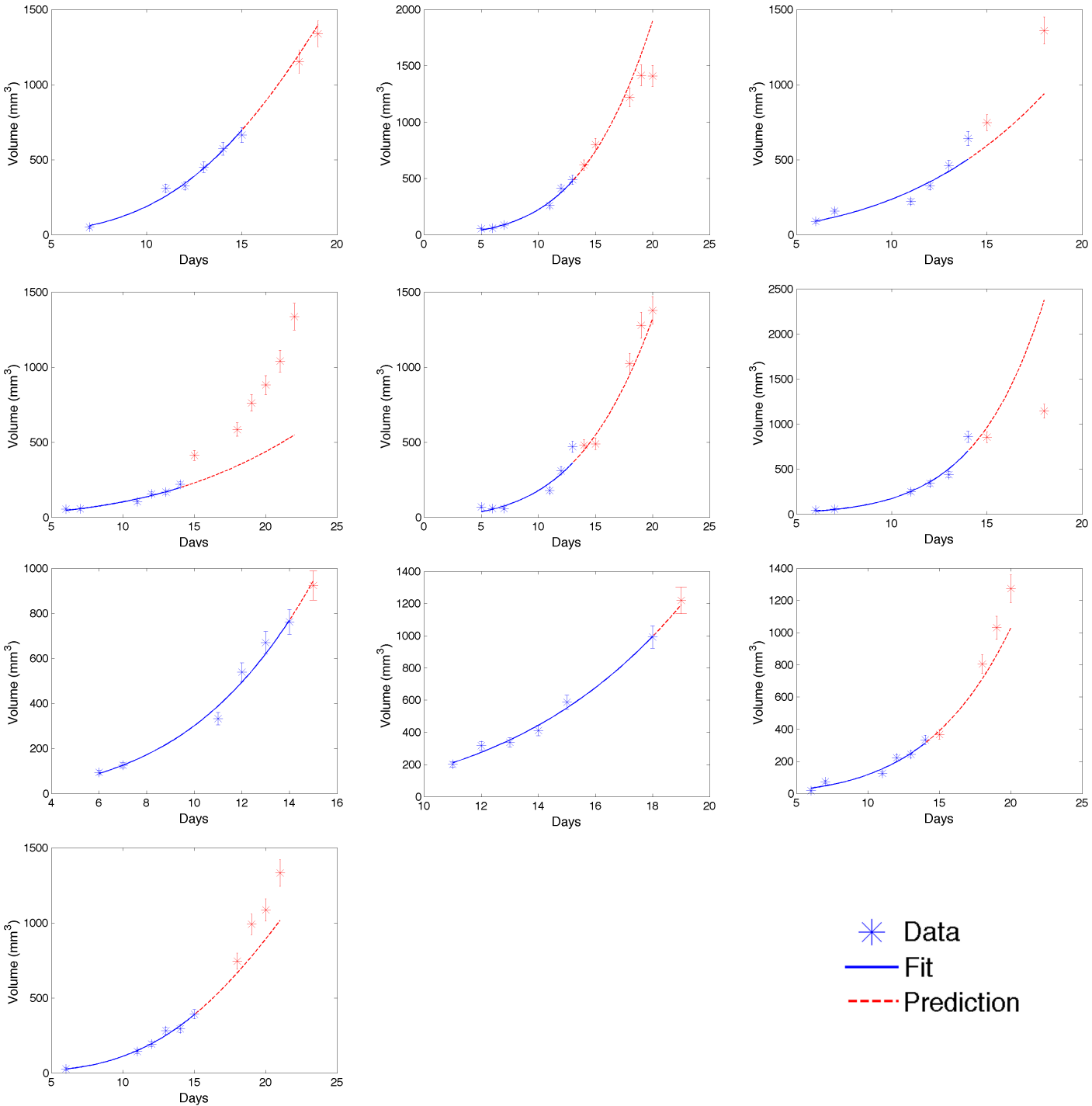

No *a priori*
A priori
A
B
Success rate (%)
PG
Exp
Gomp
Dyn CC
Volume (mm3)
Days
Data
Fit
Prediction

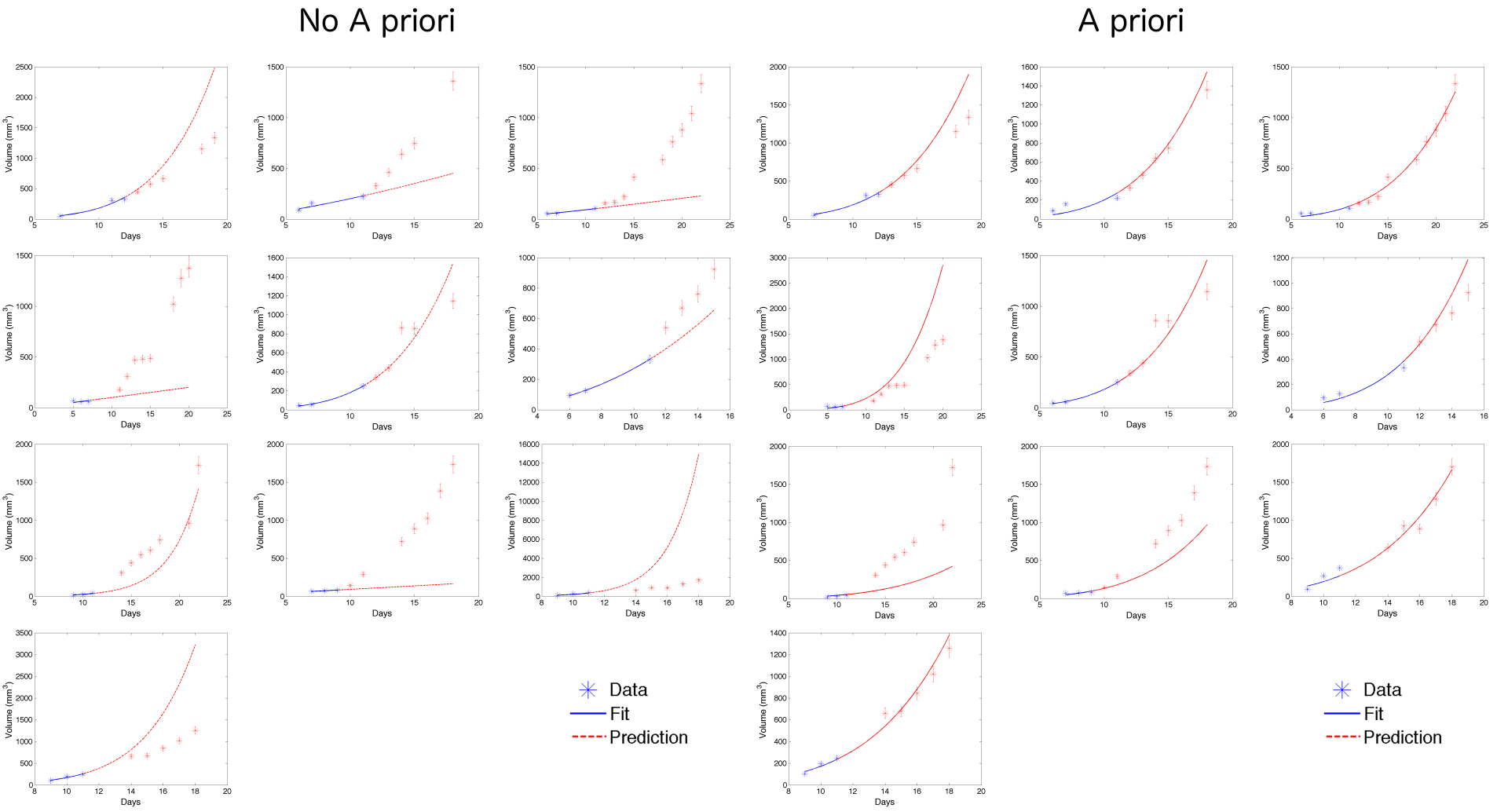
Figure 5



Supplementary Figure 1



Supplementary Figure 2



Supplementary Figure 3