HAL open science

Classical Mathematical Models for Description and Forecast of Experimental Tumor Growth

Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John M.L. Ebos, Lynn Haltky, Philip Hahnfeldt

To cite this version:

Sébastien Benzekry, Clare Lamont, Afshin Beheshti, Amanda Tracz, John M.L. Ebos, et al.. Classical Mathematical Models for Description and Forecast of Experimental Tumor Growth. 2014. hal-00922553v4

HAL Id: hal-00922553

https://inria.hal.science/hal-00922553v4

Preprint submitted on 25 Mar 2014 (v4), last revised 10 Jul 2014 (v6)

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.

Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth

Sébastien Benzekry[a,b], Clare Lamont[b], Afshin Beheshti[b], Amanda Tracz[c], John M.L. Ebos[c], Lynn Hlatky[b], Philip Hahnfeldt[b]

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

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

[c]*Department of Medicine, Roswell Park Cancer Institute, Elm & Carlton Streets, Buffalo, NY, 142631*

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

Abstract

Quantitative and discriminant analysis of a wide array of classical mathematical models was performed for description and prediction of tumor kinetics of two in vivo tumor growth experimental models, which included both ectopic syngeneic tumors (lung carcinoma) and orthotopic human xenografts (breast carcinoma). Measurement error was investigated and an adapted statistical model of the data uncertainty was derived. Combined to several goodness-of-fit criteria, it allowed assessment of the descriptive power of the models. The Gompertz and power law models were found most adequate for parsimonious description of the lung data, while linear kinetics was the main feature of the breast growth curves, best fitted by an exponential-linear model.

Study of the forecast properties of the models, using various number of data points and future prediction depths, revealed the generalized logistic and exponential-linear models as having the highest mean success rate for future growth prediction, for respectively the lung and breast data. High predictability (success rates larger than 80%) was observed in the breast setting even when using low number (3) of data points and predicting far in time (up to 12 days). In contrast, for the lung data, substantial prediction success rates (larger than 70%) could only be achieved for the next day data point. Predictability was found higher in the late phases of the growth. A method using a priori information of the parameters distribution within a learning group, for individual predictions in an independent group revealed very helpful and resulted in high improvement of the predictive power that was the most impressive for the power law model (increase of the average success rate from 14.9% to 62.7% for prediction of the global future curves using 3 data points, lung data set).

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 power of the most classical of these and identified what models are best adapted, based on data from two different experimental settings.

We also assessed the predictive power of these models and showed that the most descriptive ones were not necessarily the most predictive and had limited prediction accuracy when no statistical information is used about distribution of the models parameters in the population, for one of the experimental data set, while the other had a linear profile that allowed good predictability. When only few data points were used, we analyzed a method that takes into account the distribution of parameters in a given database for individual estimation of the models parameters. It revealed very helpful and significantly improved the prediction success rates, differentially among the models.

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 regulation of proliferation and control of the cell cycle, stroma recruitment, angiogenesis or escape from immune surveillance. In combination, these produce a macroscopic expansion of the tumor volume, thus raising the prospect of a possible general law for the global dynamics of neoplasia. 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 (injected ectopically or orthotopically), transgenic mice models or longitudinal studies of clinical images. Each scale has its own advantages and drawbacks, with increasing relevance tending to coincide with decreasing measurement precision. The data used in the current study are from two different *in vivo* experiments. The first is an experimental model of syngeneic Lewis Lung Carcinoma (LLC) subcutaneously implanted in immune-competent mice, which is a fast growing tumor that represents one of the standard for aggressiveness adopted by the National Cancer Institute in 1972 [1] such that a therapy directed at LLC, if successful, would demonstrate an efficacy of interest in human disease. A second experimental model used is a human breast cancer cell line injected orthotopically (into the mammary fat pad) of immunocompromised mice [2].

Biological study of tumor growth dynamics has already more than a 60 years history (see [3] for one of the first study devoted to the subject) and has been the matter of many investigations up to now (see [4] for an extensive review of early works and [5–8] for more recent studies). One of the most common observation from these works is that the tumor growth exhibits a slowdown in its relative growth rate [9] (or equivalently, an increase of the doubling time) both in animal [10] and human [11–13] data.

These observations raised the concern of quantification of the tumor growth curves and the possibility of general growth laws expressed by mathematical models, often written as ordinary differential equations (see [14] for a recent review). Generally, the utility of these models could be twofold: 1) testing growth hypotheses or theories by assessing their descriptive power against experimental data and 2) predicting future or past history of a tumor growth curve [10,15], as a personalized prognosis tool in a clinical context [16–20] or in order to determine the efficacy of a therapy in preclinical drug development [21,22].

Cancer modeling offers a wide range of mathematical models 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 [23]) are well-suited for studies on tumor spheroids, while the tissue scale is better described by continuous partial differential equations like reaction-diffusion models [19,24] or continuum-mechanics based models [25,26]. Here we focus on scalar data of volume longitudinal development and will only consider non-spatial models for macroscopic description of tumor kinetics, based on ordinary differential equations. A plethora of models exists at this scale [14], the most basic one being pure proliferation of a constant fraction of the tumor volume, leading to exponential growth. This model is challenged by the aforementioned observations of non-constant doubling time during tumor history, which led investigators to consider other models, the most commonly used being the Gompertz model [9,10,15,27–29]. Other examples are the logistic (used in [30,31]) or generalized logistic (as in [12]) models, models derived from balance equations of metabolic processes [32,33] such as the von Bertalanffy model [34] - which reduces to a power law (used in [5,35]) when the loss term is neglected - or a model considering an exponential growth followed by a linear regimen [36]. More recently, new models integrated tumor neo-angiogenesis in the modeling by considering a dynamical carrying capacity [37,38].

Despite several studies using specific mathematical models for description of tumor growth kinetics and a few publications comparing models performances to data of spheroid growth [31,39–41], the literature lacks comprehensive works comparing broad ranges of mathematical models for their descriptive power against *in vivo* experimental data (with the notable exception of [30]). Moreover, predictive power is very rarely considered (see [39] for an exception, about growth of tumor spheroids), although being of fundamental importance for the models' relevance and utility. The aim of the present study is to present a rational, quantitative and extensive study of the descriptive and predictive power of a broad class of mathematical models for individual *in vivo* tumor growth kinetics, based on an adapted quantification of the measurement error (and thus uncertainty) of the data. As observed by others [42], individual data should be used rather than average curves, which is the approach we used here.

The study is divided as follows. In the Materials and Methods section, we describe the experimental procedures and define the mathematical models. We also introduce the methodology for fitting the models to the data and assessing their descriptive and predictive powers. In particular, we define the goodness of fit metrics and prediction scores that were used in the sequel. Then we present the results of our analysis that consist in: 1) analysis of the measurement error and derivation of an appropriate error model subsequently used for the parameters estimation, 2) comparative descriptive power of the models against our two datasets and 3) forecasting power of the most descriptive models, with or without adjunction of *a priori* information in the estimation procedure.

Materials and Methods

Ethics statement

Animal tumor model studies were performed in strict accordance with the recommendations in the Guide for the Care and Use of Laboratory Animals of the National Institutes of Health. Protocols used were approved by the Institutional Animal Care and Use Committee (IACUC) at Tufts University School of Medicine for studies using murine lewis lung carcinoma (LLC) cells (Protocol: #P11-324) and at Roswell Park Cancer Institute (RPCI) for studies using human LM2-4^{LUC+} breast carcinoma cells (Protocol: 1227M). Institutions are AAALAC accredited and every effort 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 [43], were obtained from American Type Culture Collection (Manassas, VA). Human LM2-4^{LUC+} breast carcinoma cells are a metastatic variant originally derived from MDA-MD-231 cells and then transfected with firefly luciferase [44]. All cells were cultured in high glucose DMEM (obtained from Gibco Invitrogen Cell Culture, Carlsbad, CA or Mediatech, Manassas, VA) with 10%FBS (Gibco Invitrogen Cell Culture) and 5% CO2.

Tumor Injections

<u>Subcutaneous mouse syngeneic lung tumor model</u>
C57BL/6 male mice were used with an average lifespan of 878 days [45]. 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.

Orthotopic human xenograft breast tumor model

LM2-4^{LUC+} cells (1 × 10^6 cells) were orthotopically implanted into the right inguinal mammary fat pads of 6- to 8-week-old female severe combined immunodeficient (SCID) mice obtained from the Laboratory Animal Resource at RPCI, as previously described [2].

Tumor measurements

Tumor size was measured regularly with calipers to a maximum of 1.5 cm^3 for the lung data set and 2 cm^3 for the breast data set. 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). Volumes ranged 14 – 1492 mm^3 over time spans from 4 to 22 days for the lung model (two experiments of 10 animals each) and 202 – 1902 mm^3 over time spans from 18 to 38 days for the breast data (three experiments of 6 animals each). Plots of individual growth curves for both data sets are reported in Figure S1.

Mathematical models

For all the models, the descriptive variable is the total tumor volume, denoted by V, as a function of time t. It is assumed to be proportional to the total number of cells in the tumor. Since we dispose of the information about the number of injected cells (10^6 cells $\simeq 1$ mm^3) and to reduce the number of degrees of freedom, all the models except the exponential V_0 model have fixed initial volume condition. Although the number of cells that actually take to form a tumor is probably lower than the number of injected cells (probably around 60-80%), we considered 1 mm^3 as a reasonable approximation for $V(t = 0)$ considering that the order of magnitude of the total growth curve is 1000 mm^3.

Exponential-linear models

The simplest theory of tumor growth is assuming that all the cells are proliferating, with constant cell cycle duration T_C. This leads to exponential growth, which is also valid when not all but a constant fraction of the volume is proliferating or when the cell cycle length is a random variable with exponential distribution (assuming all the cells to be independent and identically distributed). Initial exponential phase can be assumed to be followed by a linear growth phase, as done in [36] yielding the following Cauchy problem for the volume rate of change (growth rate)

$$\begin{cases} \dfrac{dV}{dt} = a_0 V, & t \leq \tau \\ \dfrac{dV}{dt} = a_1, & t > \tau \\ V(t = 0) = V_0 \end{cases} \tag{1}$$

where coefficient a_0 is the fraction of proliferative cells times $\frac{\ln 2}{T_c}$ under the assumption of constant cell cycle length (and the fraction of proliferative cells times the inverse of the mean cell cycle length under the assumption of exponentially distributed cell cycle lengths). The coefficient a_1 drives the linear phase. Assuming that the solution of the problem (1) is continuously differentiable uniquely determines the value of τ as $\tau = \frac{1}{a_0} \log\left(\frac{a_1}{a_0 V_0}\right)$. The coefficient V_0 denotes the initial volume. From this formula, three models were considered: a)

initial volume fixed to 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 with fixed initial volume of 1 mm^3, referred to as the *exponential-linear model*.

Sigmoid models and Gompertz

A general class of models used for quantification of tumor growth kinetics have a sigmoid shape, i.e. an increasing curve with one inflexion point that asymptotically converges to a maximal volume, the carrying capacity, denoted here by K. This allows to qualitatively reproduce the experimentally observed growth slowdown [10–13] and is consistent with general patterns of organs or organisms growth. The *logistic model* is defined to have a linear decrease of the relative growth rate $\frac{1}{V}\frac{dV}{dt}$ relatively to the volume, hence yielding the formulation

$$\begin{cases} \dfrac{dV}{dt} = aV\left(1 - \dfrac{V}{K}\right) \\ V(t=0) = 1\,mm^3 \end{cases} \tag{2}$$

where a is a coefficient related to proliferation kinetics. This model can be interpreted as mutual competition between the cells (for nutrients or space, for instance), by noticing that under this model the instantaneous probability for a cell to proliferate is proportional to $1 - \frac{V}{K}$. It can be explicitly solved into

$$V(t) = \frac{V_0 K}{V_0 + (K - V_0)e^{-at}}$$

This model has been used for instance in [30]. Others [12] consider a generalization of the logistic equation defined by

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

that will be referred to as the *generalized logistic model*. In [12] the power α was fixed to $\frac{1}{4}$ but in our context we will consider it as a free parameter. When a different parameterization is employed, replacing a by $\frac{a}{\alpha}$, this model converges when $\alpha \to 0$ to the *Gompertz model*, which is defined by

$$\begin{cases} \dfrac{dV}{dt} = aV\log\left(\dfrac{K}{V}\right) \\ V(t=0) = 1\,mm^3 \end{cases} \tag{4}$$

Although first introduced in [46] for a different purpose than growth of a biological process, namely the description of human mortality for actuarial applications, the Gompertz model revealed powerful for description of growth processes in general [47] and for growth of tumors in particular, for which it was first successfully used in [27] before its applicability being confirmed on large animal data sets in [10,28] and for human breast data in [29]. The essential characteristic of the Gompertz model and fundamental difference with the logistic

growth curve is that it exhibits exponential decay of the relative growth rate. An analytic formula can be derived for the solution of (4) and is given by

$$V(t) = K\left(\frac{V_0}{K}\right)^{e^{-at}}$$

Dynamic carrying capacity

One complexity step above is a model assuming a dynamic (time-dependent) carrying capacity (CC) [37,38], assumed to represent the tumor vasculature. Considering that stimulation of the carrying capacity is proportional to the tumor surface and neglecting angiogenesis inhibition, this model is formulated in terms of two coupled equations:

$$\begin{cases} \frac{dV}{dt} = aV\log\left(\frac{K}{V}\right) \\ \frac{dK}{dt} = bV^{2/3} \\ V(t=0) = 1\,mm^3, K(t=0) = K_0 \end{cases} \tag{5}$$

and will be referred as the *dynamic CC model*. It should be noticed that this model was first developed with the intent of modeling the effect of anti-angiogenic therapies on tumor growth and not specifically for description or prediction of tumor growth. However, we believed useful to integrate it in our analysis in order to investigate and quantify whether consideration of a dynamic carrying capacity could help for these tasks.

Von Bertalanffy and power law

Von Bertalanffy [34], followed later on by others [32] proposed to derive general laws of organic growth from basic energetic principles. Stating that the net growth rate should result from the balance of synthesis and destruction, observing that metabolic rates very often follow the law of allometry (i.e., that they scale with a power of the total size) [34] and assuming that catabolic rates are in proportion of the total volume, he derived the following model for growth of biological processes

$$\frac{dV}{dt} = aV^{\gamma} - bV \tag{6}$$

Endowed with our usual assumption that $V(t=0) = 1\,mm^3$, we will refer to this model as the *von Bertalanffy model* (note that others [31] use the same name for the case $\gamma = 2/3$, which is only a particular case of the original model of [34] termed second type growth). It has already been applied for description of tumor growth in [33], although some critics were addressed to the claimed universality of the law of growth [48]. Explicit solution of the model is given by

$$V(t) = \left(\frac{a}{b} + \left(V_0^{1-\gamma} - \frac{a}{b}\right)e^{-b(1-\gamma)t}\right)^{\frac{1}{1-\gamma}}$$

From the observation that our data don't exhibit a clear saturation phase, we also considered another model derived from (6) that neglects the loss term, i.e. taking $b = 0$. This model will be termed the *power law model*. Pushing further the reasoning of [34] and arguing that the rate of synthesis of new material, in the context of tumor growth, should be proportional to

the number of proliferative cells (under the assumption of a constant cell cycle length), this model suggests that the proliferative tissue is proportional to V^{γ}. This could be further interpreted as a possibly fractional Hausdorff dimension of the proliferative tissue, when viewed as a metric subspace of the full tumor volume. For instance $\gamma = \frac{2}{3}$ could represent a proliferative rim limited to the surface of the tumor, in which case simple calculations show that the tumor radius - proportional to $V^{1/3}$ - grows linearly in time, a pattern often reported for tumor growth, for instance in the case of gliomas [18]. On the other hand, $\gamma = 1$ assumes that proliferative cells are uniformly distributed and leads to exponential growth. More generally, any power $0 < \gamma < 1$ gives a tumor growth with decreasing growth fraction (and thus decreasing relative growth rate), this fact being possibly elucidated in terms of a geometrical feature of the proliferative tissue. This model was first introduced for murine tumor growth description in [35] and was also applied in a human setting in [5].

Fit procedures and goodness of fit criteria

Individual approach

The main method we used to fit the models to the data is based on individual fits for each animal. The underlying statistical framework is to consider that the volume data y_i^j for animal j ($1 \leq j \leq J$) at time t_i^j (with $1 < i < I^j$) is a realization of a random variable Y_i^j being generated by a (deterministic) model M, which depends on a parameter vector of length P $(\beta_1^j, \dots, \beta_P^j)$, perturbed by random effects, assumed to be Gaussian. In mathematical terms:

$$Y_i^j = M(t_i^j, \beta^j) + \sigma^j E_i^j \varepsilon_i^j \tag{7}$$

where the ε_i^j are independent reduced centered Gaussian random variables and Σ_i^j is the standard deviation of the error. Statistical analysis of the measurement error was performed in order to quantify the volume dependence of the uncertainty of the data (see the Results section) and yielded the following expression

$$E_i^j = \begin{cases} (Y_i^j)^{\alpha}, & Y_i \geq V_m \\ V_m^{\alpha}, & Y_i < V_m \end{cases}$$

For a given animal j, the likelihood L^j is defined as the probability of observations $(y_1^j, \dots, y_{I^j}^j)$ under model M, parameter set β^j and expression (7), i.e. $L(\beta^j) = \mathbb{P}(y_1^j, \dots, y_{I^j}^j | \beta^j)$. Under the assumption of independent Gaussian residuals, its expression can be explicitly computed. Considering then that maximizing L is equivalent to minimize $-\ln L$ leads to the classical weighted least squares minimization problem with objective defined by

$$-\ln\left(L(\beta^j)\right) = \chi^2(\beta^j) = \sum_{i=1}^{I^j} \left(\frac{y_i^j - M(t_i^j, \beta^j)}{E_i^j} \right)^2 \tag{8}$$

Minimization was performed using the Matlab [49] function *lsqcurvefit* (trust-region algorithm), except for the generalized logistic model in the breast data setting, for which the function *fminsearch* (Nelder-Mead algorithm) was used. We refer to the supporting text S1 for more details on the numerical procedures and issues. The resulting best-fit parameter vector was denoted $\hat{\beta}^j$. Standard errors (se) of the maximum likelihood estimator, from which

confidence intervals can be derived, were used to quantify the reliability of the parameters estimated. Defining an *a posteriori* estimate of σ, denoted $NMSE$, and the weighted jacobian matrix of the model for animal j, denoted J^j and numerically estimated by *lsqcurvefit*, by

$$NMSE^j = \frac{1}{I^j - P}\chi^2(\widehat{\beta^j}), \quad J^j = \left(\frac{1}{E_i^j}\frac{\partial M}{\partial \beta_p}\left(t_i^j, \widehat{\beta_p}\right)\right)_{i,p} \tag{9}$$

normalized standard errors can be approximated, in the context of nonlinear least squares regression, by the formulas (see [50])

$$\begin{aligned} Cov^j = NMSE^j\left(J^{j^T}J^j\right)^{-1}, \quad se_p^{j,2} = \left(Cov^j\right)_{p,p}, \quad nse_p \\ = \frac{se_p}{\widehat{\beta_p}} \times 100 \end{aligned} \tag{10}$$

Different initializations of the algorithm were systematically tested to establish the practical identifiability of the models (see supporting text S2).

From the obtained $\hat{\beta}^j$, we derived various indicators of the goodness of fit. The classical Akaike Information Criterion (AIC) [51], used to compare models with different number of parameters by penalizing a large number of them, is defined by

$$AIC^j = 2\chi^2\left(\widehat{\beta^j}\right) + 2P$$

Due to the limited number of data for a given individual we rather used the AICc [51], defined by

$$AICc^j = AIC^j + \frac{2P(P+1)}{I^j - P - 1} \tag{11}$$

Another classical goodness of fit criterion that also penalizes lack of parsimony of the model is given by the Root Mean Squared Error (RMSE) and is defined by

$$RMSE^j = \sqrt{\frac{1}{I^j - P}\chi^2(\hat{\beta}^j)} \tag{12}$$

Yet another criterion is the coefficient of determination, defined by

$$R^{2,j} = 1 - \frac{\sum_{i=1}^{I^j}\left(y_i^j - M\left(t_i^j, \hat{\beta}^j\right)\right)^2}{\sum_{i=1}^{I^j}\left(y_i^j - \overline{y^j}\right)^2} \tag{13}$$

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

Eventually, we considered as an additional criterion of validity of a fit the p-value obtained from the Kolmogorov-Smirnov statistical test for normality of the residuals, these being defined by

$$Res_i^j = \frac{y_i^j - M(t_i^j, \hat{\beta}^j)}{\Sigma_i^j} \tag{14}$$

Population approach (mixed-effect models)

The procedure we explained above considers all the animals within a group to be independent. On the other hand, the mixed-effect approach, implemented in the Monolix software [52], consists in pooling all the animals together and to estimate a global distribution of the model parameters in the population. More precisely, the parameter sets $(\beta^1, \dots, \beta^J)$ are assumed to be realizations of a random variable β (here taken to be log-normally distributed). The statistical representation is then formula (7) with β instead of β^j, together with

$$\ln \beta \sim \mathcal{N}(\mu, \omega)$$

Consistently with our results on the measurement error (see the Results section), the error model (i.e. the expression of Σ) was taken to be proportional to a fixed power $\alpha = 0.84$ of the volume, with a threshold volume $V_m = 0$, because Monolix did not allow for a setting with a threshold volume. This defines a population likelihood of all animals curves, which was maximized in Monolix using a stochastic algorithm for global optimization (SAEM). From this estimation process, a population $AICc$, denoted $AICc_{pop}$ was defined, using the same formula as (11).

Models predictions methods

In the majority of the cases the general setting considered was, for a given animal j and model M, to estimate the model's parameter set using only the first n data points and to predict at a depth d, i.e. predicting the value at time $t_n^j + d$, providing that a measurement exist at this day (in which case it will be denoted $y_{n,d}^j$). The resulting best fit parameter set was denoted $\widehat{\beta_n^j}$.

Prediction metrics and success score

Goodness of a prediction was quantified using the normalized error between a model prediction and the data point under consideration, defined by

$$NE_{n,d}^j = \left| \frac{y_{n,d}^j - M(t_n^j + d, \widehat{\beta_n^j})}{\Sigma_{n,d}^j} \right| \tag{15}$$

Prediction of a single time point was considered acceptable when the normalized error was lower than three, corresponding to a model prediction within three standard deviations of the measurement error of the data and generating success results in good agreement with direct visual examinations (see Figures S2 and S3). This allowed us to define a prediction score at the level of the population as the proportion of successful predictions among all animals having measurements both at times t_n and t_{n+d} (whose set will be denoted $\mathcal{J}_{n,d}$ and total number will be denoted $J_{n,d}$). We also used the population average of the normalized error. These two metrics are formally defined by

$$S_{n,d} = \frac{\#\{NE_{n,d}^j < 3,\ j \in \mathcal{J}_{n,d}\}}{J_{n,d}}, \quad NE_{n,d} = \frac{1}{J_{n,d}} \sum_{j \in \mathcal{J}_{n,d}} NE_{n,d}^j \tag{16}$$

We derived then a global score for each model by averaging on all possible values of n and d and taking into account only situations where $J_{n,d}$ was larger or equal than a minimal number of individuals taken at 5:

$$\text{Overall mean success} = \underset{n,d|J_{n,d}\geq 5}{\text{mean}} S_{n,d} \times 100 \tag{17}$$

When assessing prediction over the total future curve, thus involving several time points, we considered the median of the normalized errors:

$$NE_{n,glob}^{j} = \text{median}\left(NE_{n,d}^{j}, 1 \leq d \leq I^{j} - n\right) \tag{18}$$

together with its associated prediction score $S_{n,glob}$ and population average $NE_{n,glob}$.

The previous metrics being dependent on our underlying measurement error, we also considered the relative error and its population average, defined by

$$RE_{n,d}^{j} = \left|\frac{y_{n,d}^{j} - M(t_n^j + d, \widehat{\beta_n^j})}{y_{n,d}^{j}}\right|, \quad RE_{n,d} = \frac{1}{J_{n,d}} \sum_{j \in \mathcal{J}_{n,d}} RE_{n,d}^{j} \tag{19}$$

A priori information

For each dataset and model, the total population was randomly and equally divided into two groups. Individual fits for the first group (the “learning” group) where performed using all the available data, generating mean values $(\overline{\beta_1}, \dots, \overline{\beta_P})$ and standard deviations $(\omega_1, \dots, \omega_P)$ of a parameter vector $(\beta_1, \dots, \beta_P)$ within the population. This information was then used when estimating the individual parameter set of a given animal from the second group (the “forecast group”), based only on a subset of its data points, by penalizing the sum of squared residuals, in the following way

$$\chi_A^2(\beta^j) = \sum_{i=1}^{n} \left(Res_i^j\right)^2 + \sum_{p=1}^{P} \left(\frac{\beta_p^j - \overline{\beta_p}}{\omega_p}\right)^2$$

with Res_i^j defined by (14). 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 sufficient to be in the convergence limit of the law of large numbers (no significant difference between 20 and 100 replicates, $p > 0.2$ by Student’s t-test). Among these simulation replicates, we only considered as significant the settings where $J_{n,d} \geq 5$. For the lung data set, this lead to no exclusion for most of the situations, the exceptions being for $S_{3,5}$ and $S_{3,6}$ where only 89/100 and 72/100 replicates were eligible, respectively. For the breast data, for depths 1 to 10, respectively 99, 16, 76, 3, 100, 3, 100, 0, 34 and 77 replicates were eligible, thus results of $S_{3,2}$, $S_{3,4}$, $S_{3,6}$, $S_{3,8}$ and $S_{3,9}$ were considered as non-significant and are not reported.

Results

Measurement error

The following method was used for quantification of the error made when measuring tumor volume with calipers, which was performed only on the lung data set. One volume per time point per cage was measured twice within a few minutes interval. This gave a total of 133 measures for which we dispose of information about the error over a wide range of volumes (20.7 – 1429 mm^3). They were analyzed by considering the following statistical representation

$$Y = y_T + \sigma E \varepsilon$$

where Y is a random variable whose realizations are the measured volumes, the true volume is denoted y_T, ε is a reduced centered Gaussian random variable and σE is the error standard deviation. The two independent measures performed, termed y_1 and y_2, were, as expected, strongly correlated (Figure 1A, $r = 0.98,\ p < 0.001$). Statistical analysis rejected variance independent of volume, i.e. constant E ($p = 0.004$, χ^2 test). A proportional error model, i.e. $E = Y$ was found only weakly significant ($p = 0.083$, χ^2 test, see 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 standard deviation, in our data. To overcome these two issues we propose the following expression

$$E = \begin{cases} Y^{\alpha}, & Y \geq V_m \\ V_m^{\alpha}, & Y < V_m \end{cases} \qquad (20)$$

In this model standard deviation of the error is proportional to a power of the volume for tumors larger than V_m while volumes smaller than V_m are considered as having the same error as V_m. The proportionality coefficient is denoted by σ. After exploration of several values of V_m and α, we found $\alpha = 0.84, V_m = 83\ mm^3$ to be able to describe dispersion of the error in our data ($p = 0.196$, χ^2 test, see Figure 1C). This yielded an empiric value of $\hat{\sigma} = 0.21$. However, for the purpose of the analysis of reliability of parameter estimates, σ was re-assessed *a posteriori* by means of (10).

This result allowed precise quantification of the measurement error inherent to our data, which was a fundamental step towards the assessment of the models' descriptive power since it allowed us to quantitatively determine whether the data we have could have been generated by the model or not.

Goodness of fit

We tested all the models for descriptive properties and quantified the goodness of fit of growth kinetics on the total population according to various criteria based on two distinct estimation procedures, one using an individual approach (minimization of weighted least squares, with weights being defined from the error model derived in the previous section, see Material and Methods) and the other one using a population approach that fits all the growth curves together. Results are reported in Figure 2 and Tables 1 and 2.

Figure 2.A depicts the representative fit of a given animal's growth curve, for each data set, using the individual approach. From visual examination, exponential 1, logistic and exponential-linear models did not seem appropriate for the lung data while exponential 1 and

logistic models did not satisfactorily fit for the breast cell line. The other models seemed to describe the growth in an adequate fashion.

These individual results were consistent with global quantifications over the total population as shown by the residuals analysis (Figure 2.C) and global goodness of fit metrics (Tables 1 and 2). For both data sets, the three-parameters models (generalized logistic (3), dynamic CC (5) and von Bertalanffy (6)) accurately fitted the data. However, they exhibited high AIC_c values, suggesting a possible over-parameterization. In the lung setting, lower $AICc$ was observed for the Gompertz and power law models, with similar values of all the criteria among them. These two models thus seemed to be the best candidates to describe the tumor growth curves of the lung data set. Consequently, they suggested a constantly decreasing relative growth rate, in accordance with the poorer descriptive power of the exponential models. In the breast setting, superior fitting power was obtained by the exponential-linear model, for all but one of the metrics considered (R^2, see Table 2). For all the animals, the fits were in the linear phase of the model, thus indicating a clear linear trend of the growth curves in the range of volumes observed. The Gompertz and power law models were also found able to describe the growth curves with higher but similar mean values of the $RMSE$ compared to the exponential-linear model and smaller maximal value of several criteria among all the animals ($\frac{1}{I}\chi^2$, $AICc$ and $RMSE$). The exponential V_0, von Bertalanffy, generalized logistic and dynamic CC models still had satisfactory descriptive power, while the logistic and exponential 1 models had a substantially higher mean squared error $\frac{1}{I}\chi^2$.

The two approaches used for the estimation process differed mostly in the way they penalized the number of parameters when computing the $AICc$, due to differences in number of data points used (one animal's time points versus all animals time points) as well as structural and numerical differences (in particular, the individual approach used a deterministic optimizer while the population approach was based on a stochastic algorithm). When using the individual parameters estimated by the Monolix software and comparing the individual relative errors generated from the fits, better goodness of fit was observed with the individual approach (see Tables S2), although fits for the highest descriptive models were visually accurate in both settings. Indeed, the population approach is better designed for settings where number of data points is too low to individually estimate the parameters, which was not our case. Both approaches clearly identified the logistic and exponential 1 models as inaccurate for description of both data sets. The six best models identified by the $AICc_{pop}$ were found to have very similar values (in the range of 50 from the model with lowest $AICc_{pop}$), indicating that this criterion was not very discriminatory. On the other hand, the $AICc$ highly penalized the models with three parameters, up to the point of ranking the logistic model before all the three parameters models in the breast data setting, despite all the other metrics indicating a lower descriptive power (Table 2). A good compromise between the two extremes of $AICc_{pop}$ and $AICc$ could be well expressed by the $RMSE$ metric (12), which is the one we selected for ranking the models in Tables 1 and 2.

Parameter values resulting from the fits are reported in Tables 3 and 4, which also show the standard errors resulting from the fit procedures. Consistently with their high $AICc$, the models with three parameters exhibited very high standard errors, indicating a poor reliability of the values of parameters estimated from the fits. This was confirmed by a study of practical identifiability performed by systematically varying the initial condition of the minimization algorithm (see supporting text S2 and Table S3). Results of inter-animal variability suggested a larger heterogeneity of growth curves in the breast data than in the lung data set, which could be explained by the different growth locations (orthotopic versus ectopic). Although not identifiable, parameter α from the generalized logistic was globally

estimated to a very low value in both data sets (Table 2), indicating a trend toward the Gompertz model, substantiated by the better $RMSE$, $AICc$ and $AICc_{pop}$ of this last model.

Taken together, our results show that, despite the complexity of internal cell populations and tissue organization, at the macroscopic scale tumor growth exhibits relatively simple dynamic that can be captured through mathematical models. Models with three parameters (dynamic CC, generalized logistic and von Bertalanffy) were found highly descriptive but not identifiable. For description of subcutaneous *in vivo* tumor growth of LLC cells, the Gompertz and power law models were found to exhibit the best compromise between number of parameters and accurate descriptive power. Orthotopic growth of LM2-4^{LUC+} cells showed a clear linear trend in the range of observed volumes, well captured by the exponential-linear, power law and Gompertz models.

Forecasting tumor growth. Individual curves

The two models that were shown unable to describe our data in the previous section, namely the exponential 1 and logistic models, were excluded from further analysis. The remaining ones were further assessed for their predictive power. The challenge considered was to estimate future growth based on parameter estimation performed on a subset of the data containing only n data points (with $n < I^j$). We refer to the Materials and Methods section for the definitions of prediction metrics and success scores.

Models' predictive power in an example setting ($n = 5$)

The first setting considered was to predict future growth based on parameters estimated using the first five data points. Figure 3 presents a representative example of predictions in this setting for a given animal of the lung data set (mouse 2, see Figure S2 for all the animals' predictions using the generalized logistic model). The success criterion that we defined in the Materials and Methods was found in agreement with direct visual examination. According to this metric, the power law, dynamic CC, von Bertalanffy and generalized logistic models seemed able to accurately predict the global future growth curve while the exponential V_0, exponential-linear and Gompertz models, although passing close to the next data point, were less accurate on the remaining of the curve, in this particular mouse.

Quantifications of the goodness of the prediction on the total population, reported in Table 5 for the lung data set (see the metric $S_{5,glob}$) showed that the prediction success depends on the mouse under consideration and that at most half of the mice could be predicted for their future global curve (using the generalized logistic model). More detailed examination of mice where prediction failed for the generalized logistic model (Figure S2) revealed that for some of them the model interpreted too strongly an initial slowdown of the growth to the point of reaching saturation too early and to a low carrying capacity, resulting in large underestimation of future data points (see mice 12, 14, 17 and 19 in Figure S2). These animals were better predicted by the power law model, which does not allow for saturation of the growth curve (results not shown). However, this was compensated by other animals better predicted by the generalized logistic model (such as mice 3 and 4), resulting overall in a higher $S_{5,glob}$ (Table 5).

Study of short term predictability, for instance at a 2 days depth (score $S_{5,2}$, Table 5) showed that no more than an average relative precision of 20% can be expected, all predictions taken together whether successful or not, the worse prediction among all mice having a relative error of at least 43% (generalized logistic model).

Consideration of the same setting for the breast data indicated substantial differences between the two settings, with larger predictability of the breast data set. Indeed an average

relative error at depth two of 13% (using the exponential-linear model) was obtained. This fact was also expressed by higher $S_{5,glob}$, although caution should be employed in this comparison since the number of points predicted included in $S_{5,glob}$ was lower in the breast setting than in the lung setting (see Figures S2 and S3). Predictions of all animals using the exponential-linear model (Figure S3) showed that this higher predictability resulted from the linear dynamics observed in the breast data.

Variable number of data points used, prediction depth and position on the curve

For evaluation of the global predictive properties of the models, settings considering a variable number n of given data points (respectively $3 \leq n \leq 9$ and $3 \leq n \leq 6$ for the lung and breast data set) and varying prediction depth d (respectively $1 \leq d \leq 7$ and $1 \leq d \leq 12$) were investigated. Results are reported in Figure 4 and Tables 5 and 6.

For the lung data set, surprisingly in view of its moderate descriptive power (Table 1) and its low identifiability (Table 3 and Tables S2), the best model for prediction was found to be the generalized logistic model, with an overall mean success rate of 58%, more than 10 points larger than the second best model (von Bertalanffy), see Table 5 and Figure 4.A. According to their similar predictive patterns in the (n, d) plane, the 4 models von Bertalanffy, Gompertz, dynamic CC and power law could be grouped together and only one of them is represented in Figure 4.A (the Gompertz model, see Figure S4.A for the predictive patterns of the other models). The exponential V_0 and exponential-linear models were found to have low predictive power (lower than 40% for the overall mean success), possibly indicating that the exponential initial phase of the growth in this data set, if real, might not be predictive of future growth. In most of the situations the prediction success was found to increase with n and decrease with d (see the Gompertz model in Figure 4.A). Situations where this did not occur (as can be seen in the generalized logistic predictive pattern of Figure 4.A) mostly resulted from underlying individuals or prediction time points not being the same. Nevertheless it was observed in some settings that increasing the number of data points was deleterious for prediction of a given future data. For instance, with mouse 19, the volumes at days 17 and 18 were successfully predicted using 4 data points (corresponding to $(n, d) = (4,3)$ and $(4,4)$) but failed with 5 data points (corresponding to $(n, d) = (5,2)$ and $(5,3)$), see Figure S4, bottom.

For the breast data set, due to the linear dynamics of the observed volumes, well captured by the exponential-linear model, predictability was found much higher than in the lung setting, with an excellent overall mean prediction success of the exponential-linear model (83.8%, see Table 6) that is, 20 points larger than the second best model (generalized logistic model). This averaged score resulted from a wide spread predictability in the (n, d) plane (Figure 4.B), with high success rates even at far future prediction depth with small number of data points used for estimation (for instance, all the 5 animals having a data point at $t_3 + 12$ were successfully predicted, $S_{3,12} = 100\%$). While the exponential V_0 showed low predictive power (a result also observed in the lung data), the von Bertalanffy, Gompertz, power law, dynamic CC and generalized logistic models had relatively good overall mean success rates, ranging from 58.8% to 63.6%. Consistently with their similar structure, globally similar predictive patterns were observed between the generalized logistic and Gompertz models (Figure S4), although a small difference in overall mean success (due to only a few animals, 4% of 34 animals being 1.36, and small differences in these settings), contrasting with the situation observed in the lung setting.

As a general results, as can be observed on the distribution of relative prediction errors (Figure 4, bottom) all the models exhibited a general trend for underestimation of predictions.

Since different growth regimen exist within the same growth curve and because in a clinically relevant setting, diagnosis might occur when the tumor is already large, we tested the predictability of the next day data point (or the second next day when using the breast data, because in this case most of the measurements were performed every two days) in two opposite situations: either using the three first available data points (scores $S_{3,1}$ and $S_{3,2}$ and relative errors $RE_{3,1}$ and $RE_{3,2}$ for the lung and breast data, respectively) or using the three first of the last four measurements, as quantified by similar metrics denoted $S_{3,1}^{f}$, $S_{3,2}^{f}$ and $RE_{3,1}^{f}, RE_{3,2}^{f}$, respectively. Volumes predicted ranged 303 $\pm$ 128 mm^3 and 909 $\pm$ 273 mm^3 in the early phase for the lung and breast data, and 1245 $\pm$ 254 mm^3 and 1383 $\pm$ 211 mm^3 in the late phases. In this last setting, in order not to artificially inject the information of the first volume being 1 mm^3 at day 0, we modified the von Bertalanffy, dynamic CC, Gompertz and power law models by fixing their initial times and volumes to the previous measurement (i.e. the fifth from the end). Interestingly, the results obtained were substantially different between the two growth phases. Better predictions were obtained when predicting the end of the curve, reaching excellent scores of 13-15/16 animals successfully predicted in the case of the LLC data (and average relative errors smaller or equal than 10%, see Table 5). Similar results were observed for the breast data (except for the generalized logistic model, which had a slightly lower $S_{3,2}^{f}$), with a 63% increase from $S_{3,2}$ to $S_{3,2}^{f}$ for the power law model and up to an 87% increase for the exponential V_0 (see the bracketed numbers in Tables 5 and 6). Hence, late phase of the growth appears more predictable, possibly because of smaller curvatures of the growth curves that lead to better identifiability of the models when using limited number of data points for estimation of the parameters, the most extreme and easiest to predict dynamics being linear growth, which was observed for the breast data and indeed found very predictable.

Overall our results suggest the generalized logistic model as best adapted for prediction of future growth curve of subcutaneous LLC cells, with substantial prediction rates (>70%) rarely reached at a future depth larger than one day. The exponential-linear model is more indicated for the orthotopic xenograft breast data with success rates larger than 70% in most of the settings and excellent scores at far future depth.

Forecasting tumor growth. A priori information.

When few number of data points are available, for example with only three, individual predictions based on individual fits only was shown to be globally limited for the lung data, especially over a large time frame (Figure 4.A, Table 5). However, this situation is likely to be the clinically relevant since few clinical examinations are performed before the beginning of therapy and usable for prognosis. On the other hand, large databases of these data might be available from previous examinations on other patients. In a preclinical setting of drug investigations, growth curves of animals from a control group could be available and usable when inferring information on the individual time course of one particular treated animal. An interesting statistical method that could potentiate these *a priori* information in these two settings consists in determining a population distribution of the models parameters, learned from a given database and to combine it with the individual parameter estimation from the available restricted data points, on a given animal. We investigated this method in order to determine if it could improve the predictive performances of the models. Each dataset was randomly divided into two groups. One group was used to learn the parameters distribution (based on the full time curves) and the other for forecast purposes (limited number of data points). For a given animal, no information from his growth curve 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. We refer to the Materials and Methods, section Models prediction methods, A priori information for more technical details. Results are reported in Figure 5.

Predictions obtained using this technique were significantly and importantly improved, going for instance from an average success score of 14.9% $\pm$ 8.35% to 62.7% $\pm$ 11.9% for prediction of the total future curve, quantified by the score $S_{3,glob}$ (see the lung data set and power law model in Figure 5.A and also Figure S5), thus allowing predictions to be made up to a far depth in the future. Indeed, 7 days future predictions could reasonably be considered (success rate of 49.2% $\pm$ 14.8%, lung data and power law model, see Figure 5.C), while their success rate was very low with direct individual prediction. At the smaller horizon of prediction of the next day data point ($S_{3,1}$), while success was moderate using an individual approach (Table 5), predictions success almost reached 100% with the method, for the lung data (power law model, Figure 5.B). Other small horizon depths also exhibited excellent prediction scores (Figure 5.C). The largest improvement of success rates for the power law model was observed for $S_{3,3}$ that went from 6.86% $\pm$ 7.47% to 75.2% $\pm$ 12.9% (more than 11 fold increase). We reported in Figure S5 the details of predictions with and without *a priori* of all animals within a given forecast group from the lung data set, when using the power law model. It can be appreciated how much addition of the information of the parameters distribution in the estimation procedure significantly improved global prediction of the growth curves. However, the impact of the addition of the a priori information was less important when using more data points for the estimation (results not shown).

Not all the models equally benefited from the addition of *a priori* information in the fitting procedure (Figure 5). Models having the lowest parameter inter-animal variability (Table 3) such as the power law, exponential-linear and exponential V_0 models, which also have better practical identifiability (Table S3) exhibited great benefit, while the three models with three parameters, including the generalized logistic model that was found to have the best predictive power in the previous setting, showed only modest benefit, or even a decrease of success rate (see $S_{3,1}$ for the von Bertalanffy model in Figure 5.B). Indeed, due to a widely spread distribution of parameters (especially K_0 for the dynamic CC model, related to the fact that dynamic carrying capacity is not observed and b for the von Bertalanffy model, due to the absence of growth saturation in our data, see Table 5), the *a priori* distribution does not contain much information for these models, and thus does not add much information to the fits. On the other hand the power law model, whose distribution in γ is particularly narrow for the lung data set (Table 3), has a much more informative *a priori* distribution that translated into the most drastic improvement of the predictive power. Despite a large inter-animal variability of parameter K, important improvement of the Gompertz model‘s prediction rate was observed, possibly due to a relatively better identifiability than the three-parameters models (Table 3).

For the breast data, due to its already high prediction score without adjunction of *a priori* information, the exponential-linear model did not benefit from the method. Similarly to the lung setting, the power law and exponential V_0 model showed important improvement, the Gompertz only moderate improvement (possibly due to a larger inter-animal variability than in the lung setting, see Table 4) and small to negative effect was observed for the three-parameters models. For prediction of the next day data point, predictability was already almost maximal without adjunction of *a priori* information and no important impact was observed.

These results demonstrate that addition of *a priori* information in the fit procedure greatly improved the forecast performances of the models, in particular when using small number of data points and low-parameterized models for data with low predictability, such as the power law model for the lung data set.

Discussion

Rigorous study of the descriptive and predictive power for a class of mathematical models for experimental tumor growth was performed.

Derivation of a specific measure error model was a fundamental consideration in the quantitative assessment of the models' performances and rejection of inaccurate growth theories. Although a proportional error model (as used by others [31]) was not strongly rejected, a more adequate error model to our data was developed. In our results, linear decay (in volume) of the relative growth rate (logistic model) had to be rejected as a non-valid theory, a result also observed by others [31] for description of experimental spheroid growth. This contrasts with results of [30] where the logistic model was found best among the models considered, which included the Gompertz model (for description of human data). This could be explained by their fit procedure, where a constant error model was used. This consideration significantly alters the results, as shown by [53] who demonstrated, using the same data set, that better fits to the same data set as [30] could be obtained with the Gompertz model, when minimizing the least square errors between the logarithms of tumor sizes and logarithms of model value (which corresponds to a proportional error model). A different error model could also generate different results on predictions. Prospective study of predictive properties on the lung data set when using a constant error model revealed changes in the ranking of the models, although the generalized logistic model was still found to be the most predictive (results not shown). However, the analysis of the impact of the error model on prediction power is beyond the scope of the present study and would warrant further investigations.

Several criteria of goodness of fit were investigated, based on two structurally different approaches for estimation of the parameters. Based on our analysis, we found the $RMSE$ to be best adapted for classification of the descriptive power of the models. This criterion, corroborated by the others, suggested the Gompertz and power law models as best adapted for description of the subcutaneous growth of LLC cells, while the kinetics of the orthotopically xenografted LM2-4^{LUC+} cells were found to be mostly linear in the range of volumes observed and well captured by exponential growth from the number of injected cells (during the unobserved phase) that switches smoothly to a linear phase (exponential-linear model). However, it should be noted that in the breast data set, no data was available during the initiation phase (below 200 mm^3) and only the linear part of a putative exponential-linear growth was observed. For the lung data, under the exponential-linear model, switching volumes between the two phases were of 258 $\pm$ 70.4 mm^3, i.e. typically of the order the minimal volumes recorded in the breast setting. Our results demonstrated that the breast data growth curves could as well be described by the power law and Gompertz models, suggesting a possible general common pattern of growth between the two experimental models and that these two theories, although possibly less accurate than others in specific phases, describe well general tumor growth curves patterns from initiation to maximal volumes of 1500 mm^3.

Despite structural similarities, important differences were noted in the parameters estimates between the two experimental models, consistently with other works emphasizing differences between ectopic and orthotopic growth [54,55]. Our results and methodology could have biological implications to help to shed light and identify more precisely the impact on kinetics of the site of implantation, although explicit comparisons cannot be made here due to the difference of cell lines used.

Our results confirmed already reported observations [9–12,28,29] that tumor growth could not be continuously exponential (constant doubling time) in the range of tumor volumes studied and consequently that it could not be explained only by proliferation of a constant fraction of the neoplasm. Exploration of the kinetics during the initial phase would warrant

further investigation, in particular for the breast model to determine to which extent growth can be approximated as linear.

The Gompertz model (exponential decay of the relative growth rate) was found to give a valid description of this growth slowdown, consistently with several previous works [9,15,28,31]. Two main critics are usually addressed to the Gompertz model. First, despite several attempts to derive the Gompertz formula from fundamental principles (well summarized in [53]), it still lacks a clear and convincing biological basis [53]. Second, as can be seen in formula (4) from the presence of the logarithm, the relative growth rate goes to infinity (or equivalently, the tumor doubling time gets arbitrarily small) when the volume becomes small, which, beyond a threshold, becomes biologically unrealistic. This consideration lead investigators [9,56] to introduce the Gomp-exp model that consists in an initial exponential phase followed by a gompertzian growth when associated doubling time becomes realistic. This approach could also be applied to any decreasing relative growth rate model but was not considered in our analysis due to the relatively large amount of cells injected and the lack of data on the initiation phase when tumors are too small to be detected and measured.

The power law model appeared as a simple, robust, descriptive and predictive mathematical model for murine tumor growth kinetics. It suggests a possible general law of macroscopic *in vivo* tumor growth: only a subset of the tumor cells proliferates, the Hausdorff measure of this subset being proportional to a constant fractional power of the volume. A hypothesis for explanation of this fractional power could be a fractal development of the tumor vasculature during tumor neo-angiogenesis [57] due to its branching nature that could generate a contact surface between blood vessels and the cancer tissue of fractional Hausdorff dimension, larger than 2 but lower than 3. The power law model showed a close match to our data and to the Gompertz curve (in the range of the observed volumes) and could reconcile the Gompertz model with the biology by giving a mechanistic explanation of the growth rate decay that naturally happens when fractional dimension of the proliferative tissue is lower than 3. However, our results were obtained in two particular settings (ectopic mouse syngeneic lung tumor and orthotopic human xenograft breast tumor models) and, although consistent with other studies that found the power law model adequate for growth of a murine mammary cell line [35] or for description of human mammography density distribution data [5], they remain to be confirmed and substantiated by extension to broader (in particular human) settings. This model should alos be taken with caution when dealing with very small volumes (at the scale of several cells for instance) for which the relative growth rate becomes very large. Indeed, the interpretation of a fractional dimension then falls, since the tumor tissue cannot be considered as a continuous medium anymore.

Practical numerical identifiability was also considered as a criterion for model comparison and revealed low identifiability of the models with three parameters such as the generalized logistic, von Bertalanffy and dynamic CC models, despite an accurate descriptive power. Consequently, we warrant caution when using these models to infer biological information from estimated parameter values, at least when the data ranges don't allow identification of last phases of saturation. It should be noticed that the dynamic CC model was not first 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. Kinetics under the influence of therapy could thus give useful additional information that could render the model identifiable in this setting.

For prediction of future growth, the models with higher predictive power were different for the two experimental settings (respectively generalized logistic model and exponential-linear for the lung and breast data sets). Surprisingly, our study revealed that models having low identifiability could nevertheless exhibit good predictive power. Indeed, on a limited time span, different parameter sets for a given model could generate the same growth curves which would thus be equally predictive. Due to its linear pattern, predictability of the breast

data was found very high (>80% success rate) up to a 12 days depth using only 3 data points. In contrast, the lung data was less predictable and could hardly be predicted to a further depth than one day, without using *a priori* information. For the Gompertz model, predictive power might be improved by using possible correlations between the two parameters of this model, as reported by others in different settings [15,56]. If a backward prediction would be considered (for instance for identification of the inception time of the tumor), we warrant the use of exponential growth for the initial, latency phase, following the lines of [9,56] (Gomp-exp model, which could also be adapted to other models such as the power law).

Adjunction of *a priori* information for improvement of the estimation procedure when few data points are used significantly improved the prediction power for the lung data set and low-parameterized models, due to narrow distributions of the parameter values, reaching almost 100% of success rate of the next day data point and high predictive performances up to a seven days depth.

Translating our results into a clinical setting raises the possibility of forecasting solid tumor growth using simple macroscopic models. Use of *a priori* information could then be a powerful method and one might think of the population distribution of parameters being learned from existing databases of previous patients examinations. However, the very strong improvement of prediction success rates that we obtained partly comes from the important homogeneity of our growth data (in particular the LLC data) that generated a narrow and very informative distribution of the parameters (for instance parameter γ of the power law model), which in turn powerfully assisted the fitting procedure. In more practical situations such as with patient data, much more heterogeneity of the growth data should be expected, which could alter the benefit of the method. For instance, in some situations (and in contrast with our data), growth could stop for arbitrarily long period of times. These dormancy periods challenge the universal applicability of a generic growth law such as the Gompertz or power law, as expressed by [48]. Description of such dormancy phenomena could be integrated by using stochastic models that would elaborate on the deterministic models reviewed here, as done in [58] for description of breast cancer growth data using the Gompertz model. Moreover, further information than just the mere tumor volume could be extracted from (functional) imaging devices, feeding more complex mathematical models that could help design more accurate *in silico* prediction tools [18,59].

Our analysis demonstrates that use of mathematical models is a valuable tool for helping preclinical anti-cancer research, as also expressed by others [21,22]. For instance, they could be used for individual assessment of drug efficacy by estimating how importantly the treated tumor deviates from its natural course and using *a priori* information learned from the control group. Another application is in their use for rational design of dose and temporal scheduling of anti-cancerous drugs [22,60,61]. Although integration of therapy remains to be added (and validated) to models such as the power law, more classical models (exponential-linear [36] or dynamic CC [38]) 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 in relevant preclinical settings.

Acknowledgments

We thank Etienne Baratchart for valuable suggestions and comments.

References

1. Talmadge JE, Singh RK, Fidler IJ, Raz A (2007) Murine models to evaluate novel and conventional therapeutic strategies for cancer. Am J Pathol 170: 793–804. doi:10.2353/ajpath.2007.060929.

2. Ebos JML, Lee CR, Cruz-munoz W, Bjarnason G a, Christensen JG, et al. (2009) Accelerated Metastasis after Short-Term Treatment with a Potent Inhibitor of Tumor Angiogenesis. Cancer Cell 15: 232–239. doi:10.1016/j.ccr.2009.01.021.

3. Collins VP, Loeffler RK, Tivey H (1956) Observations on growth rates of human tumors. Am J Roentgenol Radium Ther Nucl Med 76: Am J Roentgenol Radium Ther Nucl Med.

4. Steel GG (1977) Growth kinetics of tumours. Clarendon Press. Oxford.

5. Hart D, Shochat E, Agur Z (1998) The growth law of primary breast cancer as inferred from mammography screening trials data. Br J Cancer 78: 382–387.

6. Friberg S, Mattson S (1997) On the growth rates of human malignant tumors: implications for medical decision making. J Surg Oncol: 284–297.

7. Spratt J a S, Meyer JS (1996) Rates of growth of human neoplasms : part II. J Surg Oncol 61: 143–150.

8. Heuser L, Spratt JS, Polk HC (1979) Growth rates of primary breast cancers. Cancer 43: 1888–1894.

9. Wheldon TE (1988) Mathematical models in cancer research. Hilger Bristol.

10. Laird AK (1965) Dynamics of Tumour Growth: Comparison of Growth Rates and Extrapolation of Growth Curve To One Cell. Br J Cancer 19: 278–291.

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

12. Spratt JA, von Fournier D, Spratt JS, Weber EE (1993) Decelerating growth and human breast cancer. Cancer 71: 2013–2019.

13. Akanuma A (1978) Parameter analysis of Gompertzian function growth model in clinical tumors. Eur J Cancer 14: 681–688.

14. 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.

15. Norton L, Simon R, Brereton HD, Bogden AE (1976) Predicting the course of Gompertzian growth. Nature 264: 542–544.

16. 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.

17. 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.

18. 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.

19. 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.

20. 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.

21. Bernard A, Kimko H, Mital D, Poggesi I (2012) Mathematical modeling of tumor growth and tumor growth inhibition in oncology drug development. Expert Opin Drug Metab Toxicol 8: 1057–1069. doi:10.1517/17425255.2012.693480.

22. Simeoni M, De Nicolao G, Magni P, Rocchetti M, Poggesi I (2013) Modeling of human tumor xenografts and dose rationale in oncology. Drug Discov Today Technol 10: e365–72. doi:10.1016/j.ddtec.2012.07.004.

23. 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.

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

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

26. 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.

27. Casey AE (1934) The Experimental Alteration of Malignancy with an Homologous Mammalian Tumor Material : I . Results with Intratesticular Inoculation. Am J Cancer 21: 760–775.

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

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

30. Vaidya VG, Alexandro FJ (1982) Evaluation of some mathematical models for tumor growth. Int J Biomed Comput 13: 19–36.

31. Marusić M, Bajzer Z, Freyer JP, Vuk-Pavlović S (1994) Analysis of growth of multicellular tumour spheroids by mathematical models. Cell Prolif 27: 73–94.

32. West GB, Brown JH, Enquist BJ (2001) A general model for ontogenetic growth. Nature 413: 628–631. doi:10.1038/35098076.

33. Guiot C, Degiorgis PG, Delsanto PP, Gabriele P, Deisboeck TS (2003) Does tumor growth follow a “universal law”? J Theor Biol 225: 147–151. doi:10.1016/S0022-5193(03)00221-2.

34. Bertalanffy L Von (1957) Quantitative laws in metabolism and growth. Q Rev Biol 32: 217–231.

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

36. 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.

37. 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.

38. 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.

39. Marušić M, Vuk-Pavlović S (1993) Prediction power of mathematical models for tumor growth. J Biol Syst 1: 69–78.

40. Olea N, Villalobos M, Nuñez MI, Elvira J, Ruiz de Almodóvar JM, et al. (1994) Evaluation of the growth rate of MCF-7 breast cancer multicellular spheroids using three mathematical models. Cell Prolif 27: 213–223.

41. Wallace DI, Guo X (2013) Properties of tumor spheroid growth exhibited by simple mathematical models. Front Oncol 3: 51. doi:10.3389/fonc.2013.00051.

42. Retsky MW, Swartzendruber DE, Wardwell RH, Bame PD (1990) Is Gompertzian or exponential kinetics a valid description of individual human cancer growth? Med Hypotheses 33: 95–106.

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

44. Ebos JML, Lee CR, Bogdanovic E, Alami J, Van Slyke P, et al. (2008) Vascular endothelial growth factor-mediated decrease in plasma soluble vascular endothelial growth factor receptor-2 levels as a surrogate biomarker for tumor growth. Cancer Res 68: 521–529. doi:10.1158/0008-5472.CAN-07-3217.

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

46. Gompertz B (1825) On the Nature of the Function Expressive of the Law of Human Mortality, and on a New Mode of Determining the Value of Life Contingencies. Phil Trans R Soc B 115: 513–583. doi:10.1098/rstl.1825.0026.

47. Winsor C (1932) The Gompertz curve as a growth curve. Proc Natl Acad Sci U S A 18: 1–8.

48. Retsky M (2004) Universal law of tumor growth. J Theor Biol 229: 289. doi:10.1016/j.jtbi.2004.04.008.

49. Matlab with statistics and optimization toolboxes (2013).

50. Seber GA, Wild CJ (2003) Nonlinear regression. Wiley-Interscience.

51. Motulsky H, Christopoulos A (2004) Fitting models to biological data using linear and nonlinear regression. Oxford University Press.

52. Monolix software (2013).

53. Bajzer Ž, Vuk-Pavlović S, Huzak M (1997) Mathematical modeling of tumor growth kinetics. A Survey of Models for Tumor-Immune System Dynamics. Springer. pp. 89–133.

54. Tsuzuki Y, Carreira CM, Bockhorn M, Xu L, Jain RK, et al. (2001) Pancreas Microenvironment Promotes VEGF Expression and Tumor Growth: Novel Window Models for Pancreatic Tumor Angiogenesis and Microcirculation. Lab Invest 81: 1439–1451. doi:10.1038/labinvest.3780357.

55. Ahn KS, Jung YS, Kim J, Lee H, Yoon SS (n.d.) Behavior of murine renal carcinoma cells grown in ectopic or orthotopic sites in syngeneic mice. Tumour Biol 22: 146–153. doi:50609.

56. Demicheli R, Foroni R, Ingrosso A (1989) An exponential-Gompertzian description of LoVo cell tumor growth from in vivo and in vitro data. Cancer Res 49: 6543–6546.

57. Carmeliet P, Jain R (2000) Angiogenesis in cancer and other diseases. Nature: 249–257.

58. Speer JF, Petrosky VE, Retsky MW, Wardwell RH (1984) A stochastic numerical model of breast cancer growth that simulates clinical data. Cancer Res 44: 4124–4130.

59. 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.

60. Swierniak A, Kimmel M, Smieja J (2009) Mathematical modeling as a tool for planning anticancer therapy. Eur J Pharmacol 625: 108–121. doi:10.1016/j.ejphar.2009.08.041.

61. Barbolosi D, Freyer G, Ciccolini J, Iliadis A (2003) Optimisation de la posologie et des modalités d'administration des agents cytotoxiques à l'aide d'un modèle mathématique. B Cancer 90: 167–175.

Tables

Table 1: **Fit performances of growth models for the lung data**.

Table 2: **Fit performances of growth models for the breast data**.

Table 3: **Parameter values estimated from the fits. Lung data.**

Table 4: **Parameter values estimated from the fits. Breast data.**

Table 5: **Predictive power. Lung data.**

Table 6: **Predictive power. Breast data.**

Figures

Figure 1: **Volume measurement error**.

Figure 2: Descriptive power of the models for lung and breast data.

Figure 3: Example of predictive power.

Figure 4: Prediction.

Figure 5: *A priori* information and improvement of prediction success rates.

Supporting Information

Text S1: Numerical procedures for parameters estimation.

Text S2: Practical identifiability of the models.

Figure S1: Data.

Figure S2: Examples of individual predictions. Lung data.

Figure S3: Examples of individual predictions. Breast data.

Figure S4: Prediction.

Figure S5: Forecast improvement of the power law model when using *a priori* information and the lung data set.

Table 1: Fit performances of growth models for the lung data.

Model	$\frac{1}{I}\chi^2$	**AICc**	**AICc**$_{pop}$	**RMSE**	$\mathbf{R^2}$	**p > 0.05**	**#**
Gompertz	0.738 (0.0906 - 3.19) [3]	**2.36 (-15 - 18.4) [1]**	2108 [5]	**1.94 (0.758 - 4.42) [1]**	**0.97 (0.82 - 1)**	100	2
Power law	0.739 (0.0752 - 3.39) [4]	2.39 (-17.1 - 19) [2]	2091 [2]	1.95 (0.691 - 4.56) [2]	0.96 (0.78 - 1)	95	2
Dynamic CC	**0.646 (0.0642 - 2.9) [1]**	6.87 (-12.8 - 22.4) [4]	**2063 [1]**	1.99 (0.699 - 4.55) [3]	0.97 (0.82 - 1)	95	3
Von Bertalanffy	0.669 (0.0752 - 3.19) [2]	6.87 (-12.8 - 23.2) [5]	2096 [3]	2 (0.753 - 4.78) [4]	0.97 (0.81 - 1)	95	3
Generalized logistic	0.769 (0.125 - 2.04) [5]	9.28 (-8.83 - 19.7) [8]	2114 [6]	2.25 (0.921 - 3.94) [5]	0.96 (0.85 - 1)	100	3
Exponential V_0	1.03 (0.033 - 4.32) [6]	5.13 (-18.9 - 21.2) [3]	2099 [4]	2.32 (0.457 - 5.14) [6]	0.93 (0.68 - 1)	100	2
Exponential-linear	1.05 (0.229 - 3.62) [7]	7.28 (-2.23 - 18.7) [6]	2174 [7]	2.41 (1.28 - 4.79) [7]	0.96 (0.91 - 0.99)	100	2
Logistic	1.1 (0.237 - 3.46) [8]	7.46 (-1.04 - 18.3) [7]	2214 [8]	2.47 (1.3 - 4.68) [8]	0.96 (0.92 - 0.99)	100	2
Exponential 1	6.49 (1.47 - 11.2) [9]	20.1 (7.77 - 31) [9]	2442 [9]	5.8 (2.83 - 7.67) [9]	0.64 (0.28 - 0.94)	30	1

Models are ranked in ascending order of the RMSE, defined by expression (12). For each criterion, indicated are the mean value (among all animals) and in parenthesis the minimal and maximal values. When reported, value inside brackets is the rank of the model for the goodness-of-fit metric. The model ranking first is highlighted in bold. $1/I\chi^2$ is the minimal value of the objective that was minimized in the individual fits approach (variance of the weighted residuals, defined in (8)), $AICc$ is defined in (11), $AICc_pop$ is the $AICc$ resulting from the mixed-effect estimation (see Materials and Methods) and R^2 is defined in (13). Values reported in the p column are percentages of animals were Kolmogorov-Smirnov test for normality of residuals was not rejected at the significance level of 0.05. # = number of parameters. J = 20 animals.

Table 2: Fit performances of growth models for the breast data.

Model	$\frac{1}{I}\chi^2$	**AICc**	**AICc**$_{pop}$	**RMSE**	$\mathbf{R^2}$	**p > 0.05**	**#**
Exponential-linear	**0.438 (0.0755 - 2.33) [1]**	**2.31 (-10 - 14.9) [1]**	**2832 [1]**	**1.6 (0.774 - 3.95) [1]**	0.92 (0.66 - 0.99)	100	2
Gompertz	0.465 (0.0698 - 1.56) [4]	2.72 (-12.9 - 13) [2]	2866 [3]	1.66 (0.666 - 3.23) [2]	0.92 (0.67 - 0.99)	97	2
Power law	0.485 (0.0759 - 1.54) [5]	3.04 (-6.79 - 13.8) [3]	2913 [7]	1.7 (0.776 - 3.37) [3]	0.92 (0.61 - 0.99)	100	2
Exponential V_0	0.564 (0.0507 - 1.76) [7]	4.12 (-5 - 14.8) [4]	2870 [4]	1.83 (0.634 - 3.74) [4]	0.9 (0.56 - 0.99)	100	2
Von Bertalanffy	0.443 (0.0717 - 1.54) [2]	12.9 (-7.08 - 33) [6]	2876 [6]	1.84 (0.739 - 3.81) [5]	**0.93 (0.67 - 0.99)**	97	3
Generalized logistic	0.464 (0.0673 - 1.6) [3]	13.3 (-7.59 - 32.9) [7]	2870 [5]	1.9 (0.716 - 3.79) [6]	0.92 (0.68 - 0.99)	97	3
Dynamic CC	0.524 (0.0855 - 2.39) [6]	14 (-1.19 - 33.6) [8]	2862 [2]	2.01 (1.01 - 4.47) [7]	0.91 (0.55 - 0.99)	100	3
Logistic	0.692 (0.0175 - 1.98) [8]	5.88 (-8.24 - 13.8) [5]	2921 [8]	2.04 (0.372 - 3.64) [8]	0.86 (0.65 - 0.99)	100	2
Exponential 1	10.4 (2.94 - 16.4) [9]	19.9 (12.3 - 26.6) [9]	3518 [9]	7.58 (4.04 - 9.54) [9]	-0.91 (-5.9 - 0.88)	24	1

Models are ranked in ascending order of the RMSE, defined by expression (12). For each criterion, indicated are the mean value (among all animals) and in parenthesis the minimal and maximal values. When reported, value inside brackets is the rank of the model for the goodness-of-fit metric. The model ranking first is highlighted in bold. $1/I\chi^2$ is the minimal value of the objective that was minimized in the individual fits approach (variance of the weighted residuals, defined in (8)), $AICc$ is defined in (11), $AICc_pop$ is the $AICc$ resulting from the mixed-effect estimation (see Materials and Methods) and R^2 is defined in (13). Values reported in the p column are percentages of animals were Kolmogorov-Smirnov test for normality of residuals was not rejected at the significance level of 0.05. # = number of parameters. J = 34 animals.

Table 3 Lung data

Model	Par.	Unit	Median value (CV)	Mean normalized std error (CV)
Power law	a	$\left[mm^{3(1-\gamma)} \cdot day^{-1}\right]$	0.921 (38.9)	11.9 (48.7)
	γ	-	0.788 (9.41)	4 (53.4)
Gompertz	a	$\left[day^{-1}\right]$	0.0792 (42.3)	13.7 (65.3)
	K	$\left[mm^3\right]$	12194 (248)	83 (99.6)
Exponential-linear	a_0	$\left[day^{-1}\right]$	0.49 (19.3)	3.08 (41.5)
	a_1	$\left[mm^3 \cdot day^{-1}\right]$	115.6 (22.6)	15.7 (40.7)
Dynamic CC	a	$\left[day^{-1}\right]$	0.399 (101)	351 (112)
	b	$\left[mm^{-2} \cdot day^{-1}\right]$	2.65 (53.1)	136 (88.8)
	K_0	$\left[mm^3\right]$	2.61 (134)	3.72e+05 (410)
Von Bertalanffy	a	$\left[mm^{3(1-\gamma)} \cdot day^{-1}\right]$	1.53 (48.9)	725 (217)
	γ	-	0.909 (11.4)	45.3 (75.1)
	b	$\left[day^{-1}\right]$	0.376 (114)	1.54e+08 (203)
Generalized logistic	a	$\left[day^{-1}\right]$	13.4 (77.1)	1.08e+04 (116)
	K	$\left[mm^3\right]$	9081 (19.3)	568 (71)
	α	-	0.0055 (84.5)	1.11e+04 (114)
Exponential V_0	V_0	$\left[mm^3\right]$	13.2 (47.9)	28.9 (55)
	a	$\left[day^{-1}\right]$	0.257 (15.4)	7.49 (48.3)
Logistic	a	$\left[day^{-1}\right]$	0.502 (17.5)	3.03 (48.9)
	K	$\left[mm^3\right]$	1297 (23.1)	17.2 (43.8)
Exponential 1	a	$\left[day^{-1}\right]$	0.399 (13.8)	2.87 (24.5)

Shown are the median values within the population and in parenthesis the coefficient of variation (CV), expressed in percent and defined as the standard deviation within the population divided by mean and multiplied by 100) that quantifies inter-animal variability. Last column represents the normalized standard errors (nse) of the maximum likelihood estimator, defined in (10).

Table 4 Breast data

Model	Par.	Unit	Median value (CV)	Mean normalized std error (CV)
Power law	a	$\left[mm^{3(1-\gamma)} \cdot day^{-1}\right]$	1.32 (74.1)	31.2 (48.6)
	γ	-	0.58 (23)	12.1 (62.2)
Gompertz	a	$\left[day^{-1}\right]$	0.0719 (26.4)	12.5 (65.4)
	K	$\left[mm^3\right]$	2988 (526)	49.2 (165)
Exponential-linear	a_0	$\left[day^{-1}\right]$	0.31 (16.8)	6.22 (65.9)
	a_1	$\left[mm^3 \cdot day^{-1}\right]$	67.8 (33.2)	12.9 (45.4)
Dynamic CC	a	$\left[day^{-1}\right]$	2.63 (84.1)	584 (337)
	b	$\left[mm^{-2} \cdot day^{-1}\right]$	0.871 (44.3)	67.7 (261)
	K_0	$\left[mm^3\right]$	9.84 (134)	4.32e+03 (420)
Von Bertalanffy	a	$\left[mm^{3(1-\gamma)} \cdot day^{-1}\right]$	1.26 (33.9)	497 (103)
	γ	-	0.887 (19.3)	138 (60.3)
	b	$\left[day^{-1}\right]$	0.321 (91.3)	1.48e+08 (300)
Generalized logistic	a	$\left[day^{-1}\right]$	6.13 (36)	-
	K	$\left[mm^3\right]$	3029 (111)	-
	α	-	0.0106 (35.4)	-
Exponential V_0	V_0	$\left[mm^3\right]$	68.2 (57.2)	34.5 (50.8)
	a	$\left[day^{-1}\right]$	0.0846 (27.7)	13.7 (44.1)
Logistic	a	$\left[day^{-1}\right]$	0.305 (10.2)	3.17 (34.9)
	K	$\left[mm^3\right]$	1221 (31.4)	11.8 (73.8)
Exponential 1	a	$\left[day^{-1}\right]$	0.223 (5.9)	3.72 (21.3)

Shown are the median values within the population and in parenthesis the coefficient of variation (CV), expressed in percent and defined as the standard deviation within the population divided by mean and multiplied by 100) that quantifies inter-animal variability. Last column represents the normalized standard errors (nse) of the maximum likelihood estimator, defined in (10). They are not reported for the generalized logistic model in the breast setting because in this case estimation was performed using fminsearch, which does not generate approximate value of the objective Jacobian.

Table 5: Lung data

Model	**Overall mean success**	$\mathbf{S_{5,glob}}$	$\mathbf{S_{5,2}}$	$\mathbf{RE_{5,2}}$	$\mathbf{S_{3,1}}$	$\mathbf{RE_{3,1}}$	$\mathbf{S^f_{3,1}}$	$\mathbf{RE^f_{3,1}}$
Generalized logistic	58.0	10/20	6/11	0.20 (0.004 - 0.43)	10/14	0.16 (0.001 - 0.55)	14/16 [22]	0.10 (0.003 - 0.31)
Von Bertalanffy	44.7	8/20	6/11	0.20 (0.019 - 0.51)	8/14	0.29 (0.010 - 1.08)	15/16 [64]	0.10 (0.007 - 0.26)
Gompertz	43.6	9/20	8/11	0.23 (0.064 - 0.57)	7/14	0.26 (0.030 - 0.81)	15/16 [88]	0.10 (0.001 - 0.30)
Dynamic CC	42.2	7/20	7/11	0.20 (0.010 - 0.48)	7/14	0.27 (0.002 - 1.01)	13/16 [62]	0.14 (0.012 - 0.65)
Power law	42.0	7/20	6/11	0.21 (0.019 - 0.52)	8/14	0.29 (0.010 - 1.08)	15/16 [64]	0.08 (0.007 - 0.29)
Exponential V_0	39.3	6/20	6/11	0.33 (0.039 - 1.48)	9/14	0.31 (0.035 - 1.20)	15/16 [46]	0.09 (0.000 - 0.27)
Exponential linear	36.0	8/20	5/11	0.22 (0.037 - 0.43)	10/14	0.21 (0.029 - 0.99)	15/16 [31]	0.10 (0.010 - 0.33)

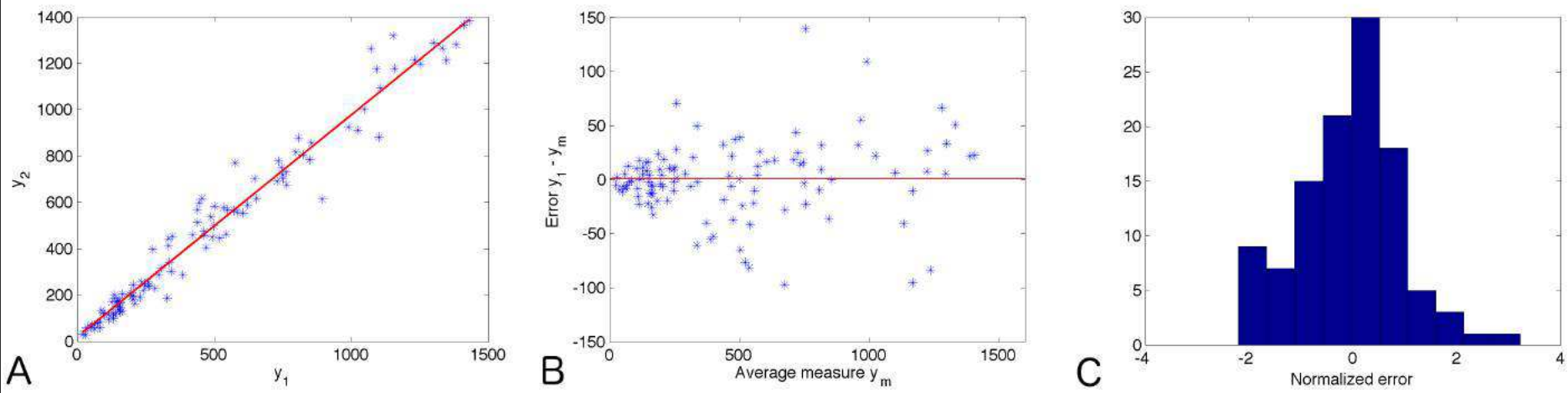
Models are presented in descending order of overall mean success (defined in (17)). $S_{n,d}$, defined in (16), is the success score for prediction when using n data points and predicting at future depth d, i.e. time $t_n + d$ (see Materials and Methods). For relative errors (19), mean value among animals is reported with ranges in parenthesis. $S^f_{3,1}$ and $RE^f_{3,1}$ stand for the success rates and relative errors for predictions of the late phase (see text for details). Reported in brackets in the $S^f_{3,1}$ column are the percent increase between $S_{3,1}$ and $S^f_{3,1}$.

Table 6: Breast data

Model	Overall mean success	$S_{5,glob}$	$S_{5,2}$	$RE_{5,2}$	$S_{3,2}$	$RE_{3,2}$	$S^f_{3,2}$	$RE^f_{3,2}$
Exponential linear	83.8	20/25	17/23	0.13 (0.014 - 0.36)	4/7	0.27 (0.079 - 0.84)	17/20 [49]	0.10 (0.001 - 0.32)
Generalized logistic	63.6	19/25	17/23	0.14 (0.003 - 0.40)	5/7	0.18 (0.037 - 0.44)	14/20 [-2]	0.15 (0.005 - 0.45)
Dynamic CC	63.3	20/25	19/23	0.14 (0.013 - 0.53)	4/7	0.28 (0.071 - 0.86)	13/20 [14]	0.18 (0.001 - 0.43)
Power law	62.3	18/25	17/23	0.15 (0.028 - 0.57)	3/7	0.33 (0.092 - 0.97)	14/20 [63]	0.14 (0.001 - 0.41)
Gompertz	59.6	18/25	17/23	0.15 (0.001 - 0.54)	4/7	0.24 (0.076 - 0.59)	14/20 [22]	0.14 (0.002 - 0.42)
Von Bertalanffy	58.8	18/25	17/23	0.15 (0.008 - 0.54)	3/7	0.31 (0.078 - 0.87)	14/20 [63]	0.14 (0.007 - 0.43)
Exponential V0	47.7	14/25	17/23	0.16 (0.003 - 0.66)	3/7	0.37 (0.065 - 1.24)	16/20 [87]	0.13 (0.024 - 0.37)

Models are presented in descending order of overall mean success (defined in (17)). $S_{n,d}$, defined in (16), is the success score for prediction when using n data points and predicting at future depth d, i.e. time $t_n + d$ (see Materials and Methods). For relative errors (19), mean value among animals is reported with ranges in parenthesis. $S^f_{3,2}$ and $RE^f_{3,2}$ stand for the success rates and relative errors for predictions of the late phase (see text for details). Reported in brackets in the $S^f_{3,2}$ column are the percent increase between $S_{3,2}$ and $S^f_{3,2}$.

Figure 1: **Volume measurement error**. A. First measured volume against second one . Also plotted is the regression line (correlation coefficient r = 0.98, slope of the regression = 0.96). B. Error $y_1 - y_m$ against approximation of the volume given by the average of the two measurement $y_m = (y_1+y_2)/2$. The χ^2 test rejected Gaussian distribution of constant variance (p = 0.004) C. Histogram of the normalized error applying the error model given by (20) with $\alpha = 0.84$ and $V_m = 83\ mm^3$. It shows Gaussian distribution (p = 0.196, χ^2 test) with standard deviation $\sigma = 0.21$.



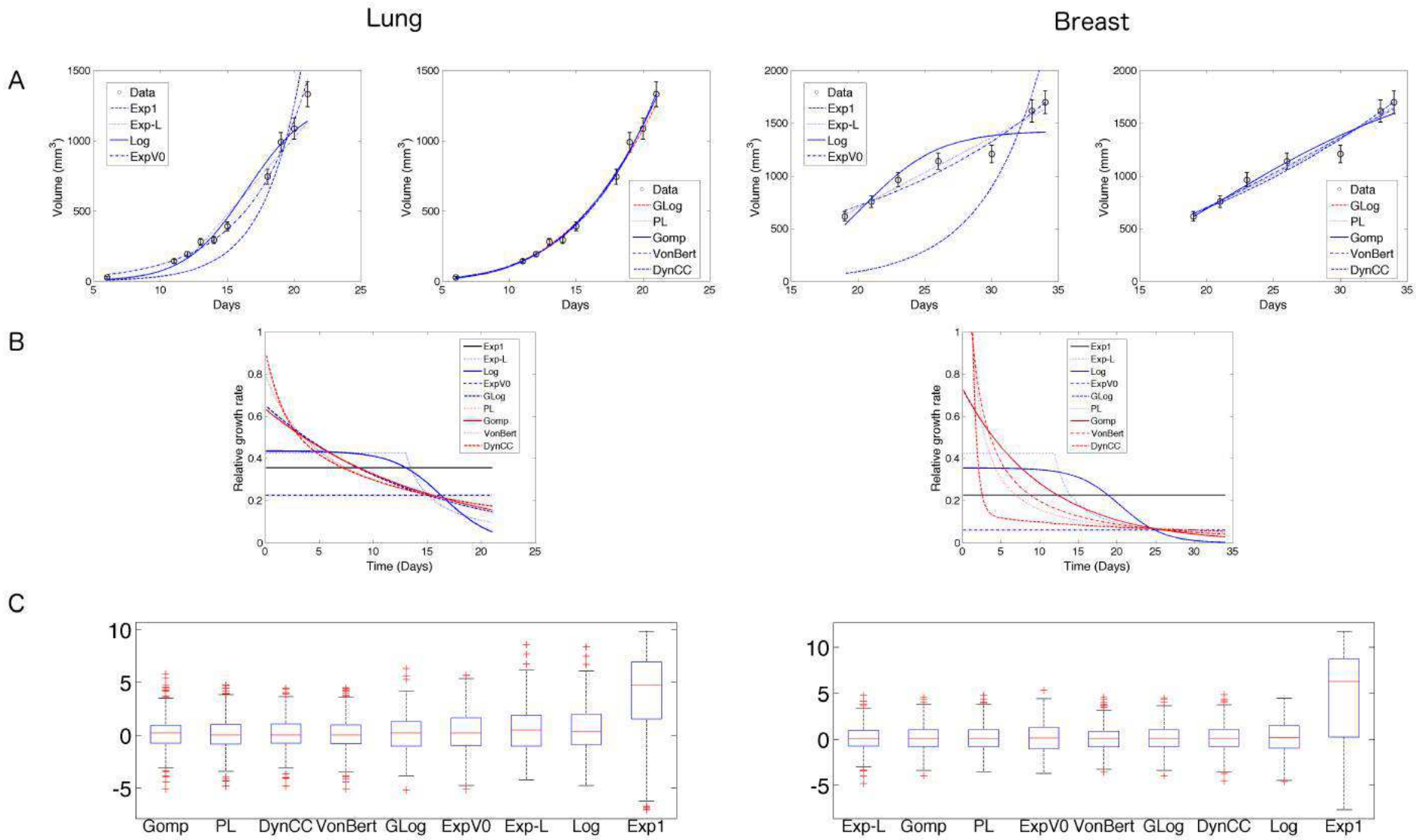


Figure 2: Descriptive power of the models for lung and breast data. A. Representative examples of all growth models fitting the same growth curve (animal 10 for lung, animal 14 for breast). Errorbars correspond to the standard deviation of the *a priori* estimate of measurement error. In the lung setting, curves of the Gompertz, power law, dynamic CC and von Bertalanffy models are visually indistiguinshable. For the breast example, this occurs for the curves of the generalized logistic and Gompertz models. B. Corresponding relative growth rate curves. Curves for von Bertalanffy and power law are identical in the lung setting C. Residuals distributions, in ascending order of mean *RMSE* (12) over all animals. Residuals (see formula (14) for their definition) include fits over all the animals and all the time points. Exp1 = exponential 1, Exp-L = exponential-linear, Exp = exponential , Log = logistic, GLog = generalized logistic, PL = power law, Gomp= Gompertz, VonBert = von Bertalanffy, DynCC = dynamic CC.

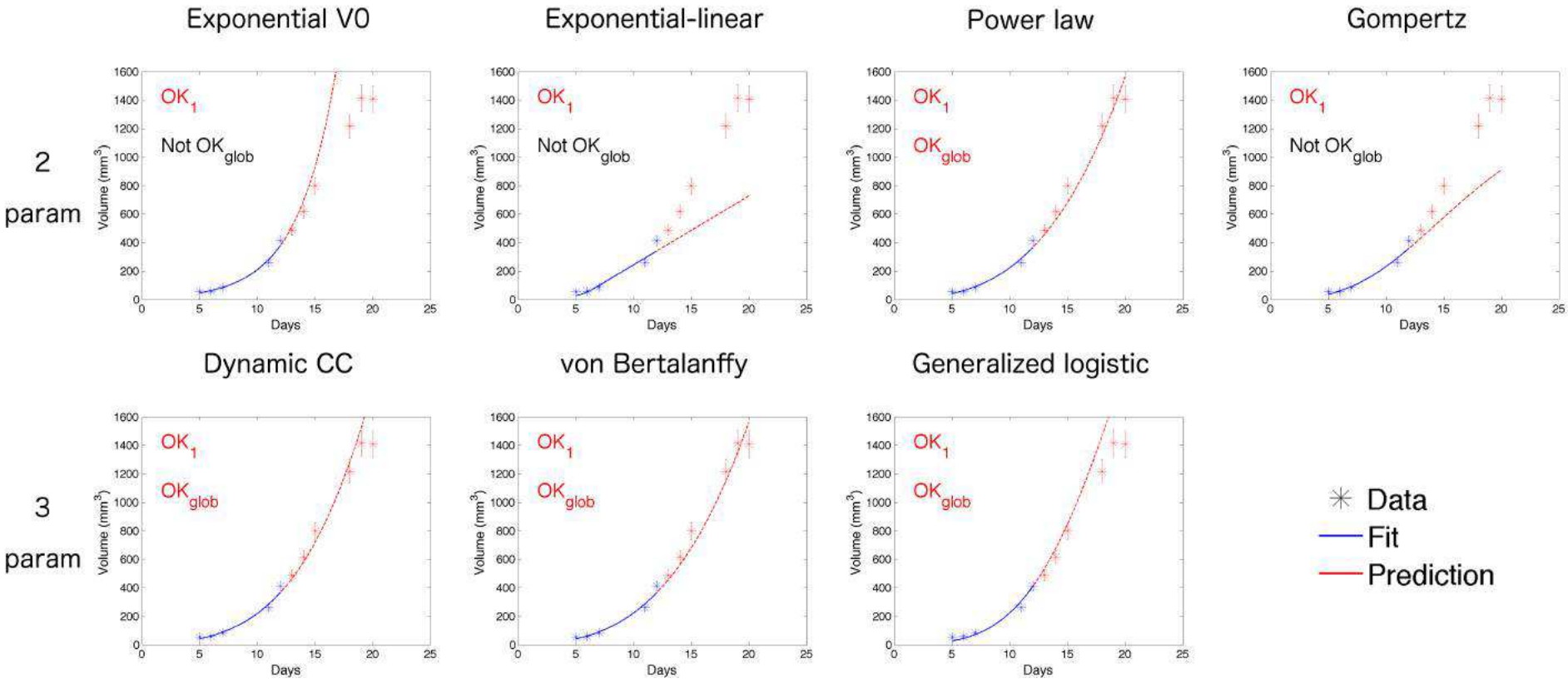


Figure 3: Example of predictive power. Representative example of the forecast performances of the seven best descriptive models for the lung data set (animal number 2). Five data points were used to estimate the animal parameters and predict future growth. Prediction success of the models are reported for the next day data point (OK_1) or global future curve (OK_{glob}), based on the criterion of a normalized error smaller than 3 (meaning that the median model prediction is within 3 standard deviations of the measurement error) for OK_1 and the median of this metric over the future curve for OK_{glob} (see Materials and Methods for details).

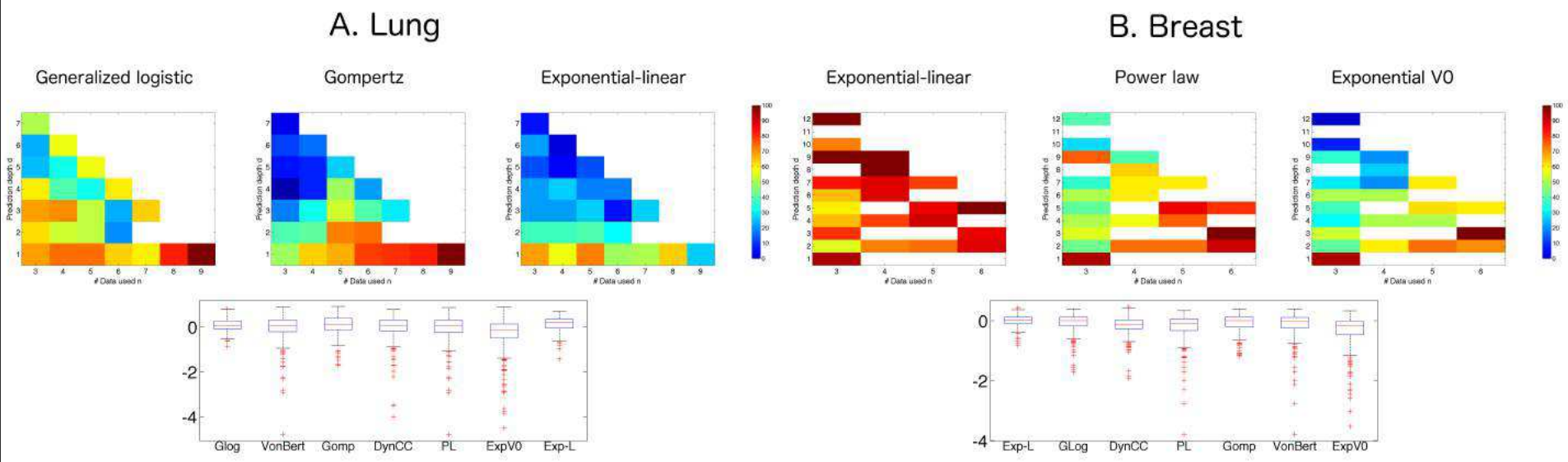


Figure 4: Prediction. Test of the predictive power of some representative models (high, medium and low predictive power) depending on the number of data points used for estimation of the parameters (*n*) and the prediction depth in the future (*d*). Top: at position the color represents percentage of successfully predicted animals when using data points and forecasting the time point t_n+d, i.e. the score $S_{n,d}$ (multiplied by 100), defined in (16). This proportion only includes animals having measurements at these two time points, thus values at different raw on the same column or reverse might represent predictions in different animals. White squares correspond to situations where this number was too low (<5) and thus success score, considered not significative, was not reported. Bottom: distribution of the relative error of prediction, all animals and settings pooled together. Models are ordered in ascending order of mean success score reported in Tables 5 and 6. A. Lung data. B. Breast data.

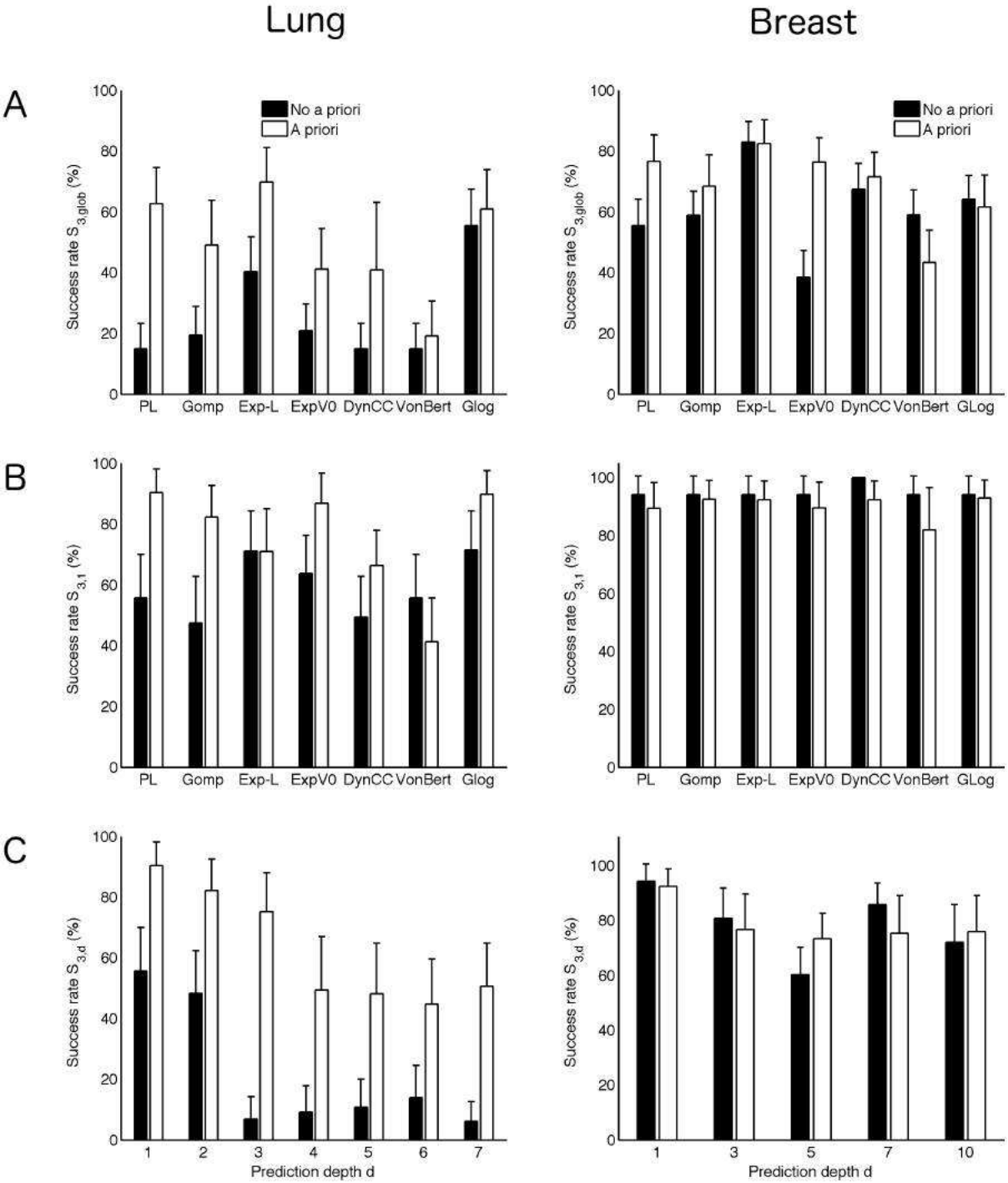


Figure 5: *A priori* information and improvement of prediction success rates. A. Success rates of the models over all simulation replicates (mean ± standard deviation) when using *a priori* information on the parameters population distributions, *n = 3* data points and predicting the global future growth curve, quantified by the score $S_{3,glob}$. Differences were statistically significant in all situations (p<0.005 by Student's t-test, except for the generalized logistic in the breast setting, p<0.05), except for the breast data set and exponential-linear model (p=0.65). B. Benefit of the method for prediction of the next day, using three data points (score $S_{3,1}$). Lung: differences were significant (p<0.005) for all but the exponential-linear model. Breast: differences were significant (p<0.005) for all but the Gompertz model (p=0.074), the exponential-linear model (p=0.058) and the generalized logistic model (p=0.18) C. Prediction improvement at various prediction depths, using the power law model (lung data) or the exponential-linear model (breast data). Lung ; all differences were significant (p<0.001). Breast : differences were not significant (p>0.05) for *d=1* (p=0.058) and *d=10* (p=0.073). Results of $S_{3,2}$, $S_{3,4}$, $S_{3,6}$, $S_{3,8}$ and $S_{3,9}$ were not considered significant (see Materials and Methods) and are not reported.