Supplementary text for

Interplay between distribution of live cells and growth dynamics of solid tumours

E. Milotti, V. Vyshemirsky, M. Sega and R. Chignola

Contents

List of Figures

1 Details of model development

In this section we complete some of the calculations presented only in summary form in the main paper.

1.1 Dynamics of live cell volume

In the main text we introduce the dynamics of total cell volume

$$\frac{dV}{dt} = \alpha V_a(t) - \delta\left(V(t) - V_a(t)\right) = \alpha F(t)V(t) - \delta V(t)\left[1 - F(t)\right] \tag{S.1}$$

and at the same time we relate total cell volume V to the volume of live cells V_a by means of equation

$$V_a(t) = F(t)V(t) \tag{S.2}$$

Then

$$\begin{aligned}
\frac{dV_a}{dt} &= F(t)\frac{dV}{dt} + V(t)\frac{dF}{dt} \\
&= F(t)\left[\alpha V_a(t) - \delta\left(\frac{V_a(t)}{F(t)} - V_a(t)\right)\right] + \frac{V_a(t)}{F(t)}\frac{dF}{dt} \\
&= \left[\alpha F(t) + \delta + \frac{d\ln F}{dt}\right]V_a(t) - \frac{\delta}{F(t)}V_a(t)
\end{aligned} \tag{S.3}$$

Equation (S.3) corresponds to equation (2) in the main text.

1.2 Live cell fraction

In the main text we consider the depth-dependent fractional density of live cells $f(s)$, which is well approximated by the exponential function:

$$f(s) = \exp\left(-s/\lambda\right) \tag{S.4}$$

Using equation (S.4) we find that the total volume taken by live cells in a spherical tumour cluster is

$$V_a(r) = \int_0^r 4\pi s^2 \exp\left(-\frac{r-s}{\lambda}\right) ds \tag{S.5}$$

where r is the cluster radius. Integrating equation (S.5) with standard elementary methods, we find

$$\begin{aligned}
V_a &= 4\pi\left[(r^2\lambda - 2r\lambda^2 + 2\lambda^3) - 2\lambda^3\exp\left(-\frac{r}{\lambda}\right)\right] \\
&= \left(\frac{4\pi}{3}r^3\right)\left[3\left(\frac{\lambda}{r} - 2\frac{\lambda^2}{r^2} + 2\frac{\lambda^3}{r^3}\right) - 6\frac{\lambda^3}{r^3}\exp\left(-\frac{r}{\lambda}\right)\right]
\end{aligned} \tag{S.6}$$

and the fraction F of live cells in this nearly spherical tumour-cell cluster is

$$F = \frac{V_a}{V} = \left[3\left(\frac{\lambda}{r} - 2\frac{\lambda^2}{r^2} + 2\frac{\lambda^3}{r^3}\right) - 6\frac{\lambda^3}{r^3}\exp\left(-\frac{r}{\lambda}\right)\right] \tag{S.7}$$

where, obviously, the cluster radius is a function of time, $r = r(t)$.

A series expansion of the exponential in (S.7) leads to the following series for $F(t)$:

$$F(r) = 6 \sum_{n=0}^{\infty} \frac{1}{(n+3)!} \left(\frac{r}{\lambda}\right)^n \tag{S.8}$$

and we see that $\lim_{\lambda \to \infty} F(r) = 1$ (i.e., in the limit of a very large λ all cells are alive), and that $\lim_{r \to 0} F(r) = 1$ for all λ's.
Moreover, for very large radii, $F(r) \sim 3\lambda/r$: this is a rather obvious result if we think that in very large tumour spheroids, the live volume is limited to a thin shell of nearly constant thickness $\sim \lambda$, and volume $V_a \sim 4\pi r^2 \lambda$. Thus the fraction of live cells is $F(r) \sim 4\pi r^2 \lambda / (4\pi r^3/3) = 3\lambda/r$ for very large radii.

1.3 Approximate form of the total live cell fraction $F(r)$

If we assume a fixed, size-independent decay constant λ, then the leading term for large radii in equation (S.7) is $3\lambda/r$. Such an asymptotic behaviour, as well as the limit $\lim_{r \to 0} F(r) = 1$, can also be obtained with a much simpler F-function:

$$F = \frac{3\lambda}{3\lambda + r} \tag{S.9}$$

so that

$$V_a(r) = \frac{4\pi r^3}{3} \frac{3\lambda}{3\lambda + r} \tag{S.10}$$

Figure S1 compares the "exact" live-cell fraction (S.7) with the approximate function (S.9), which seems to be a reasonable approximation in a biological context.

Using the approximate live-cell fraction (S.9), the complete evolution equations are

$$\frac{dV}{dt} = (\alpha + \delta) \frac{3\lambda}{3\lambda + r(t)} V(t) - \delta V(t) \tag{S.11}$$

for the total cluster volume and

$$\frac{dr}{dt} = \frac{r(t)}{3} \left[(\alpha + \delta) \frac{3\lambda}{3\lambda + r(t)} - \delta \right] \tag{S.12}$$

for the cluster radius. The asymptotic (saturation) radius r_∞ corresponds to a vanishing derivative in equation (S.12), and this yields immediately

$$r_\infty = \frac{3\alpha\lambda}{\delta} \tag{S.13}$$

The corresponding asymptotic volume is

$$V_\infty = V(r_\infty) = \frac{4\pi r_\infty^3}{3} = \frac{36\pi\alpha^3\lambda^3}{\delta^3} \tag{S.14}$$

and the asymptotic total fraction of live cells is

$$F_\infty = F(r_\infty) = \frac{3\lambda}{r_\infty + 3\lambda} = \frac{\delta}{\alpha + \delta} \tag{S.15}$$

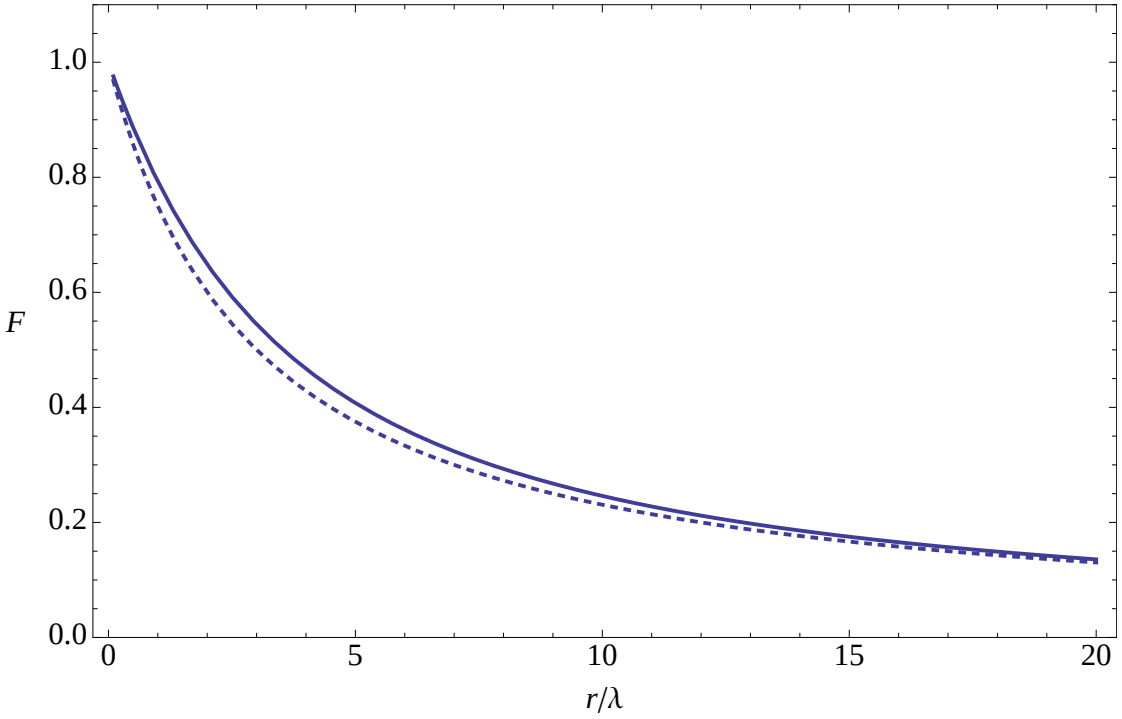

Figure S1: Plot of the fraction of live cells F vs. r/λ in the exponential model eq. (S.7) (solid line), and its approximation, eq. (S.9) (dotted line). The maximum difference between these expressions is less than 5%.

1.4 Size-dependent λ

In the foregoing sections we assumed a constant λ, however the computer simulations indicate that λ is weakly size dependent (see figure 1 in the main text):

$$\lambda(r) = \lambda_0 + \lambda_1 \exp\left(-r/\zeta\right) \tag{S.16}$$

With this variable λ, all the derivations at constant r are left unchanged, while $F(r)$ becomes

$$F(r) = \frac{3\lambda(r)}{r + 3\lambda(r)} \tag{S.17}$$

This expression has the same asymptotic behaviour as the simpler expression (S.9) (see figure S2).

1.5 The model recast as a differential system

As explained in the main text, the model can be recast as a differential system. The differential equation for volume can be written in the following form:

$$\frac{dV}{dt} = \left[(\alpha + \delta)\, F(t) - \delta\right] V(t) \tag{S.18}$$

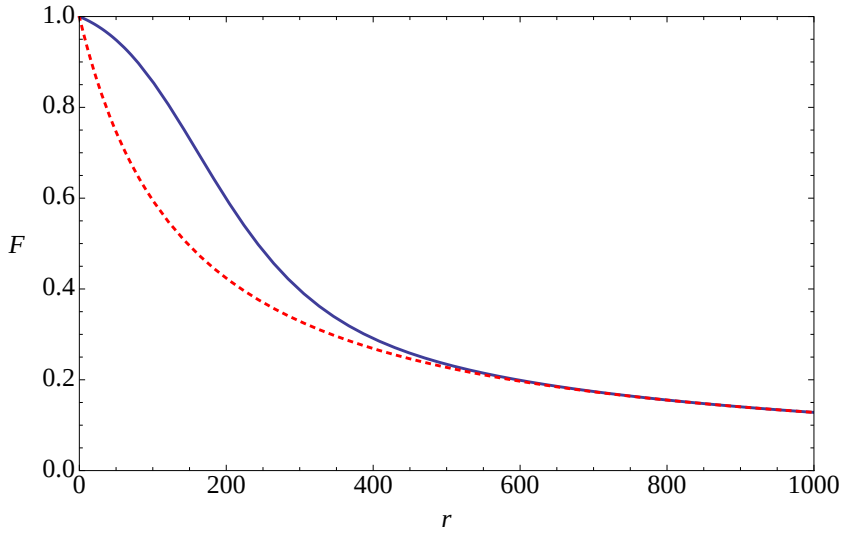

Figure S2: Plot of the fraction of live cells F vs. r (μm), with variable (solid, blue line) $\lambda(r)$ and fixed (dotted, red line) λ_c. Here the $\lambda(r)$ is the same as in Fig. 1C in the main text (i.e., $\lambda(r) = \lambda_0 + \lambda_1 \exp(-r/\zeta)$, with $\lambda_0 = 49$ μm, $\lambda_1 = 444$ μm, and $\zeta = 92$ μm), and $\lambda_c = \lambda_0 = 49$ μm.

and therefore we can also write

$$\frac{dV}{dt} = \gamma(t) V(t) \tag{S.19}$$

where the γ function is defined as follows

$$\gamma(t) = (\alpha + \delta)\, F(t) - \delta = (\alpha + \delta)\, \frac{V_a(t)}{V(t)} - \delta \tag{S.20}$$

so that $\gamma(0) = \alpha$. Then, taking the derivative of expression (S.20), we find

$$\frac{d\gamma}{dt} = (\alpha + \delta)\, \frac{dF}{dt} = (\alpha + \delta) \left(\frac{1}{V(t)} \frac{dV_a}{dt} - \frac{V_a(t)}{V^2(t)} \frac{dV}{dt} \right) = \frac{(\alpha + \delta)}{V(t)} \left(\frac{dV_a}{dt} - \frac{V_a(t)}{V(t)} \frac{dV}{dt} \right) \tag{S.21}$$

In the case of the exponentially decaying live-cell fraction (S.4), we can evaluate the time derivative of V_a from

$$\begin{aligned} V_a(t + \Delta t) &= \int_0^{r(t+\Delta t)} 4\pi s^2 \exp\left(-\frac{r(t+\Delta t) - s}{\lambda} \right) ds \\ &\approx \int_0^{r(t)} + \int_{r(t)}^{r(t) + \frac{dr}{dt}\Delta t} 4\pi s^2 \exp\left(-\frac{r(t) - s}{\lambda} \right) \left(1 - \frac{1}{\lambda} \frac{dr}{dt} \Delta t \right) ds \\ &\approx V_a(t) + 4\pi r^2(t) \frac{dr}{dt} \Delta t - \frac{1}{\lambda} \frac{dr}{dt} \Delta t V_a(t) \end{aligned} \tag{S.22}$$

i.e.,

$$\frac{dV_a}{dt} = \left[4\pi r^2(t) - \frac{1}{\lambda}V_a(t)\right]\frac{dr}{dt} = \frac{dV}{dt} - \frac{1}{\lambda}V_a(t)\frac{dr}{dt} = \gamma(t)V(t) - \frac{r}{3\lambda}V_a(t)\gamma(t) \tag{S.23}$$

and finally

$$\begin{aligned}\frac{d\gamma}{dt} &= \frac{(\alpha+\delta)}{V(t)}\left[\gamma(t)V(t) - \frac{r(t)}{3\lambda}V_a(t)\gamma(t) - \frac{V_a(t)}{V(t)}\gamma(t)V(t)\right] \\ &= -(\alpha+\delta)\left[\left(\frac{r(t)}{3\lambda}+1\right)F\left(r(t)\right) - 1\right]\gamma(t)\end{aligned} \tag{S.24}$$

We already know that $\lim_{r\to 0} F(r) = 1$, and that $\lim_{r\to\infty} F(r) = 0$, and we notice that

$$\lim_{r\to\infty}\frac{r}{3\lambda}F(r) = 1 \tag{S.25}$$

then the term in square brackets in expression (S.24)

$$\left(\frac{r(t)}{3\lambda}+1\right)F(t) - 1 \tag{S.26}$$

vanishes both at $r = 0$ and $r \to \infty$. An analytical study of this expression that utilizes expression (S.7) involves transcendental equations: however, the resulting expression depends on the single scale factor λ and therefore we obtain a fairly complete information on it simply plotting it vs. r/λ (see figure S3). The equation for the position of the maximum can be solved numerically, and it yields $r_0 \approx 3.63231\ \lambda$.

To summarize: the present model can be written as a differential system

$$\frac{dV}{dt} = \gamma(t)V(t) \tag{S.27a}$$

$$\frac{d\gamma}{dt} = -(\alpha+\delta)\left[\left(\frac{r(t)}{3\lambda}+1\right)F(t) - 1\right]\gamma(t) \tag{S.27b}$$

with $\gamma(0) = \alpha$.

1.6 The differential system that describes the Gompertz model

The Gompertz model is defined by the single equation

$$V(t) = V(0)\exp\left[\frac{\alpha_G}{\beta_G}\left(1 - e^{-\beta_G t}\right)\right] \tag{S.28}$$

which can be recast in the form of a differential system

$$\frac{dV}{dt} = \gamma_G(t)V(t) \tag{S.29a}$$

$$\frac{d\gamma_G}{dt} = -\beta_G\gamma_G(t) \tag{S.29b}$$

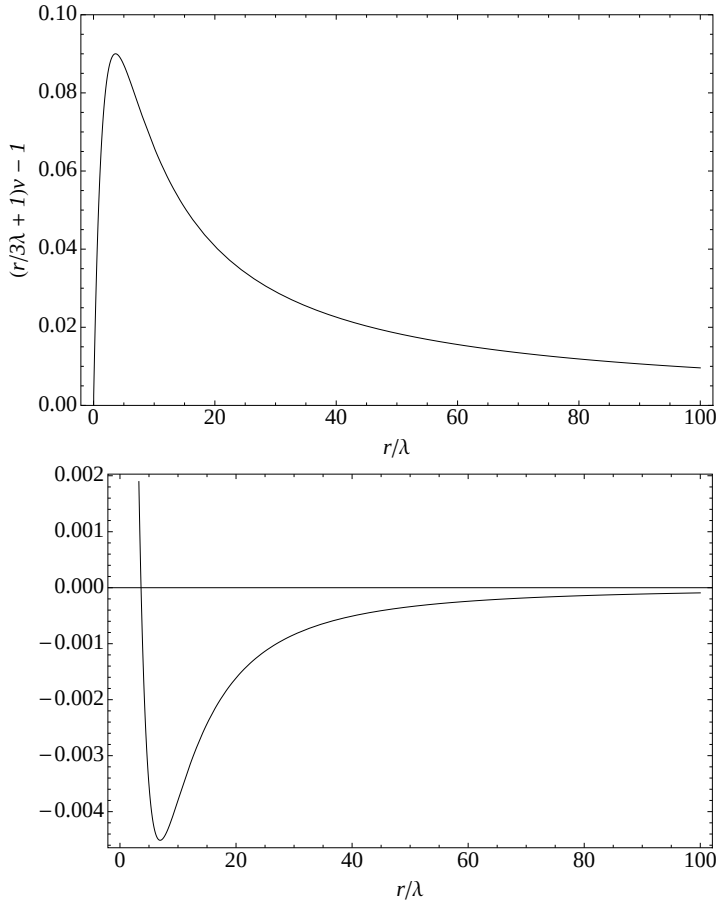

Figure S3: Upper panel: plot of the expression (S.26). The maximum is at $r_0 \approx 3.63231\ \lambda$. Notice the wide range r/λ: the variation of the expression is comparatively slow, the maximum derivative is in the origin. Lower panel: plot of the derivative of expression (S.26).

with $\gamma_G(t) = \alpha_G e^{-\beta_G t}$, and $\alpha_G = \gamma_G(0)$.

From the equation for the volume we can also find a corresponding equation for r:

$$\frac{dr}{dt} = \frac{r}{3}\gamma_G(t) \tag{S.30}$$

therefore r also follows a Gompertz law, but with a modified exponent.

Comparing equations (S.24) and (S.29b), we see that the Gompertz model is a sort of approximation of the new model, where we replace the slowly varying expression $(\alpha + \delta)\left[(r(t)/3\lambda + 1)\, F\left(r(t)\right) - 1\right]$ in (S.24) with the constant β_G in (S.29b).

2 Bayesian methods

In this work we have performed data analysis at several levels: in the case of the individual exponential estimates (figure 1A in the main text) we have used straightforward least-squares fits. Elsewhere we have used the Bayesian inference framework for model assessment and evidential model comparison. In this section we briefly review Bayesian inference and the Bayes factors, the basic tool used to rank models (we refer the reader to the text by Bernardo and Smith (1994) for more details).

2.1 Bayesian inference

Formally, Bayesian inference is statistical inference in which evidence or observations are used to update or to infer the probability that a hypothesis may be true. To perform such inference we need to define a way to express our initial beliefs and describe the process by which some evidence or observations can be used to update these beliefs.

Applying Bayesian inference methods requires the formal representation of the available knowledge. This should include the statistical model for the problem, and *a priori* information about the model parameters, as we assume that the statistical model is parametric.

In the cases when we have several competing hypotheses about some phenomenon, and therefore several competing models of it, we also associate an *a priori* probability $p(M_i)$ to each model, which describes the degree of initial belief that a particular model is the most appropriate one to describe the observed phenomenon.

Our initial beliefs (initial state of information) about the values of parameters of each available statistical model of the system are, most often, uncertain and therefore distributed according to some probability density function $p(\theta_i|M_i)$. This probability distribution is called *"a prior distribution of model parameters"*.

When some new information D about the modelled phenomenon is acquired, we update our beliefs according to Bayes' theorem. The updated distribution of our beliefs is called *"a posterior distribution of model parameters"*. D can correspond to the data from a newly performed experiment, or new information published in a recent paper. Bayes' theorem defines how the posterior can be obtained from the prior, generally:

$$p(\theta_i|M_i, D) = \frac{p(D|M_i, \theta_i) \cdot p(\theta_i|M_i)}{\int p(D|M_i, \theta_i) \cdot p(\theta_i|M_i) d\theta_i}. \tag{S.31}$$

Here the probability $p(D|M_i, \theta_i)$ to produce data D with model M_i given parameters θ_i is called *"likelihood"* (see, for example, Cox and Hinkley, 1974; Gelman et al., 1995).

In this paper we consider two complex models defined using non-linear ordinary differential equations. In cases such as these, it is not possible to perform inference analytically due to the complexity of the integrals involved, and we need some numerical methods to be able to evaluate the posteriors, such as the Monte Carlo methods (Robert and Casella, 2004; Gilks et al., 1995). We employed the Sequential Monte Carlo (SMC) sampler proposed by Del Moral et al. (2006) to find parameter posteriors for our models. This sampler employs a strategy of sequential importance sampling starting with an easy

to sample prior distribution and eventually converging to the desired posterior, through a sequence of artificial intermediate distributions (Neal, 2001).

2.2 Model Comparison and Bayes Factors

The methodology presented in this section allows one to rank competing hypotheses by the evidential support from experimental data, and therefore evaluate relative confidence values for such hypotheses. A complete comprehensive overview of Bayes factors and model comparison can be found, e.g., in Kass and Raftery (1995).

In the cases when a discrete set of competing hypotheses is considered, the hypotheses can be ranked by the ratio of their posterior probabilities. For a pair of hypotheses H_1 and H_2 represented with models M_1 and M_2 the ratio is

$$\frac{p\left(M_1|D\right)}{p\left(M_2|D\right)}. \tag{S.32}$$

Taking a prior distribution of beliefs in preference of each hypotheses π into account, and in the case when hypotheses are represented by parametric models, this ratio is:

$$\frac{p(M_1|D)}{p(M_2|D)} = \frac{\pi\left(M_1\right)}{\pi\left(M_2\right)} \times \frac{p\left(D|M_1\right)}{p\left(D|M_2\right)} = \frac{\pi\left(M_1\right)}{\pi\left(M_2\right)} \times \frac{\int p(D|M_1,\theta_1)\cdot p(\theta_1|M_1)d\theta_1}{\int p(D|M_2,\theta_2)\cdot p(\theta_2|M_2)d\theta_2} \tag{S.33}$$

The ratio of the marginal likelihoods for two competing hypotheses:

$$\frac{\int p(D|M_1,\theta_1)\cdot p(\theta_1|M_1)d\theta_1}{\int p(D|M_2,\theta_2)\cdot p(\theta_2|M_2)d\theta_2} \tag{S.34}$$

is called *the Bayes factor*.

Bayes factors are used to test competing hypotheses, and update corresponding beliefs using formula (S.33).

When using models with continuous parameter space the problem becomes quite complex, as Bayes factors have to be evaluated by integration. In the vast majority of practical problems these integrals cannot be evaluated analytically, and therefore numerical methods are required to estimate them. These integrals are called marginal likelihoods, and we give a brief overview of some numerical methods to estimate them in the next section (see Section 2.3).

The Bayes factor is a summary of the evidence provided by the data in favor of one hypothesis, represented by a model, as opposed to another. Jeffreys (1961) suggested interpreting Bayes factors in half-units on the $\log_{10}$ scale. Pooling two of his categories together for simplification we display his scale in Table S1.

These categories are not a calibration of the Bayes factor, as it already provides a meaningful interpretation as probability, but rather a rough descriptive statement about standards of evidence in scientific investigation.

Kass and Raftery (1995) propose a slight modification to this scale, and use natural logarithms instead. This modified scale is shown in Table S2.

There are a number of publications on the controversy between Bayesian and non-Bayesian testing procedures. The following four issues are usually considered:

$\log_{10}(B)$	B	Evidence support
0 to 1/2	1 to 3.2	Not worth more than a bare mention
1/2 to 1	3.2 to 10	Substantial
1 to 2	10 to 100	Strong
> 2	> 100	Decisive

Table S1: Interpretation of the Bayes factor as evidence support categories according to Jeffreys (1961)

$2\ln(B)$	B	Evidence support
0 to 2	1 to 3	Not worth more than a bare mention
2 to 6	3 to 20	Positive
6 to 10	20 to 150	Strong
> 10	> 150	Very strong

Table S2: Interpretation of the Bayes factor as evidence support categories according to Kass and Raftery (1995)

1. P values used in non-Bayesian significance testing are not similar to the posterior probability that the null hypotheses is correct. Jeffreys (1961) considers this problem and discusses the results obtained with both approaches.

2. Non-Bayesian tests tend to reject null hypotheses in very large samples, whereas Bayes factors do not. This has been a problem in sociology, where the data sets can contain thousands of cases. Facing this problem, sociologists have taken to ignoring significance tests and using other criteria and informal methods when comparing models. An example with $n = 113,566$ samples was discussed by Raftery (1986), where a meaningful model that explained 99.7% of the deviance was rejected by a standard chi-squared test with a P value of about 10^{-120} but was nevertheless favored by the Bayes factor. Bayes factors are now widely used in sociology, usually with BIC (Bayesian Information Criterion) as an approximation.

3. Bayes factors can be applied to both nested[1] and non-nested models, while application of non-Bayesian significance tests to non-nested models is difficult. This problem is briefly discussed in Kass and Raftery (1995).

4. Non-Bayesian significance tests were designed for comparison of two models, but practical data analysis often involves more than two models, at least implicitly. In such a case, performing multiple significance tests to guide a search for the best model can give very misleading results (e.g. Freedman, 1983). This problem can be avoided by taking model uncertainty into account and employing Bayes factors

[1]Nested models are statistical models with model parameters arranged in a hierarchical structure.

(e.g. Raftery et al., 1993).

Arkinson (1978) has noted some examples when Bayes factors favored the simpler model H_0 even when a more complex model H_1 was correct. Smith and Spiegelhalter (1980) demonstrated that this occurs only when the models are so close that there is almost no loss in predictive power when cutting back to the simpler model, so that Bayes factors can be considered as *a fully automatic Occam's razor*[2].

In our paper we rely on the hypotheses testing results obtained with Bayes factors. However, computing such Bayes factors is a challenging problem, as the marginal likelihoods for nonlinear models have to be evaluated to obtain these. In the following section we discuss alternative methods for estimation of the marginal likelihoods.

2.3 Estimation of the Marginal Likelihoods

Evaluation of marginal likelihoods is required to perform hypotheses testing and model comparison with Bayes factors. A review of different methods for evaluating marginal likelihoods can be found in Newton and Raftery (1994), Kass and Raftery (1995), and Chib (1995).

The main problem is that the marginal likelihood

$$p(D|M) = \int p(D|M,\theta) \cdot p(\theta|M) d\theta \tag{S.35}$$

can be evaluated *analytically* only in very special cases, e.g. when the likelihood belongs to the exponential family, and conjugate priors are used. The models considered in our paper are based on nonlinear ordinary differential equations that contribute to the likelihood. In such cases analytical integration of the marginal likelihood is impossible. Brute force *numerical integration* can rarely be applied as it quickly becomes computationally intractable. This leaves us with the only practical option of considering methods for approximate evaluation of marginal likelihoods. We use *thermodynamic integration* or *path sampling* methods (Ogata, 1989; Gelman, 1998) for approximate estimation of marginal likelihoods.

The method of thermodynamic integration originates in Statistical Physics (for an overview see Neal, 1993), where the marginal likelihood is equivalent to the so-called partition function and its logarithm to the free energy. The computations required to perform thermodynamic integration are still quite intensive, but the results are usually more stable (Gelman, 1998).

This method is based on the following principles: suppose that there are two unnormalized distributions $q_0(\theta)$ and $q_1(\theta)$, defined on the same parameter space Θ. We can normalize these densities dividing them by normalization constants.

$$p_i(\theta) = \frac{1}{Z_i} q_i(\theta),\ i = 0, 1, \tag{S.36}$$

[2]Occam's razor is a principle which states that the explanation of any phenomenon should make as few assumptions as possible. Thus, the simplest model which explains the evidence sufficiently should be chosen as the most appropriate one.

where

$$Z_i = \int_\Theta q_i(\theta) d\theta,\ i = 0, 1 \tag{S.37}$$

To perform the evaluation of log-ratio

$$\mu = \ln\left(\frac{Z_1}{Z_0}\right) = \ln Z_1 - \ln Z_0 \tag{S.38}$$

a continuous and differentiable path $(q_\beta)_{0\leq\beta\leq1}$ can be defined in the space of unnormalized densities, joining q_0 and q_1. Similarly,

$$p_\beta(\theta) = \frac{1}{Z_\beta} q_\beta(\theta), \tag{S.39}$$

where

$$Z_\beta = \int_\Theta q_\beta(\theta) d\theta. \tag{S.40}$$

Taking the derivative of $\ln Z_\beta$ with respect to β:

$$\begin{aligned}\frac{\partial \ln Z_\beta}{\partial \beta} = \frac{1}{Z_\beta}\frac{\partial Z_\beta}{\partial \beta} &= \frac{1}{Z_\beta}\frac{\partial}{\partial \beta}\int_\Theta q_\beta(\theta) d\theta \\ &= \frac{1}{Z_\beta}\int_\Theta \frac{\partial q_\beta(\theta)}{\partial \beta} d\theta = \int_\Theta \frac{1}{q_\beta(\theta)}\frac{\partial q_\beta(\theta)}{\partial \beta}\frac{q_\beta(\theta)}{Z_\beta} d\theta \\ &= \int_\Theta \frac{\partial \ln q_\beta(\theta)}{\partial \beta} p_\beta(\theta) d\theta = E_{p_\beta(\theta)}\left[\frac{\partial \ln q_\beta(\theta)}{\partial \beta}\right],\end{aligned} \tag{S.41}$$

where $E_{p_\beta(\theta)}[\cdots]$ is the expectation with respect to $p_\beta(\theta)$. Defining the *potential*

$$\mathcal{U}(\theta) = \frac{\partial \ln q_\beta(\theta)}{\partial \beta}, \tag{S.42}$$

we obtain

$$\frac{\partial \ln Z_\beta}{\partial \beta} = E_{p_\beta(\theta)}\left[\mathcal{U}\right]. \tag{S.43}$$

Integrating over $[0, 1]$ yields the log-ratio μ:

$$\mu = \ln Z_1 - \ln Z_0 = \int_0^1 \frac{\partial \ln Z_\beta}{\partial \beta} d\beta = \int_0^1 E_{p_\beta(\theta)}\left[\mathcal{U}\right] d\beta. \tag{S.44}$$

To compute this integral, a Sequential Monte Carlo sampler is usually run over a sequence of bridging distributions p_β. Expectations of the potential can then be estimated as averages on these samples. This computation is performed for a series of values of β between 0 and 1.

The log-ratio μ can then be estimated by numerical integration using trapezoidal (as in Friel and Pettitt, 2006) or Simpson's scheme (as in Lartillot and Philippe, 2006).

Assuming that $q_0(\theta)$ above is the prior $p(\theta|M)$, and $q_1(\theta)$ is the unnormalized posterior $p(D|M,\theta)p(\theta|M)$, and the corresponding normalization constants are $Z_0 = 1$ (as the prior is already normalized) and $Z_1 = p(D|M)$, the resulting log-ratio μ is the logarithm of the marginal likelihood.

Defining $q_\beta(\theta)$ as a path in the probability densities space which connects the prior and the posterior:

$$q_\beta(\theta) = p(D|M,\theta)^\beta p(\theta|M), \tag{S.45}$$

the potential takes a simple form:

$$\mathcal{U}(\theta) = \frac{\partial \ln q_\beta(\theta)}{\partial \beta} = \ln p(D|M,\theta). \tag{S.46}$$

Then, the logarithm of the marginal likelihood we are seeking an estimate for is

$$\ln p(D|M) = \mu = \ln Z_1 - \ln Z_0 = \int_0^1 E_{p_\beta(\theta)} \left[\ln p(D|M,\theta)\right] d\beta. \tag{S.47}$$

There are a number of ways to select a schedule for β to estimate this integral. In our paper, we use the schedule proposed by Friel and Pettitt (2006), and select these values as

$$\beta_i = a_i^c,\ a_i = \frac{i}{N},\ i = 0, \ldots, N. \tag{S.48}$$

Good results can usually be achieved with $N \in [20, 100]$ and $c = 3$ or $c = 4$.

3 Results of the Bayesian data analyses

Here we summarize the results of the Bayesian data analyses with a set of figures, which are referred to in the main text.

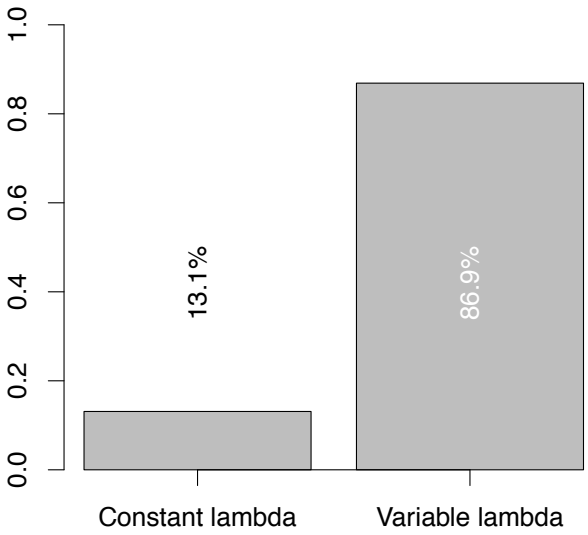

Figure S4: The bar graph shows the *a posteriori* probabilities of the of the constant-lambda and of the variable-lambda version of the new model, and it shows that previous data are not sufficient to discriminate between the two versions of the new model, although there is slightly stronger support for the model with variable-lambda. This analysis corresponds to the two curves shown in figure 2D.

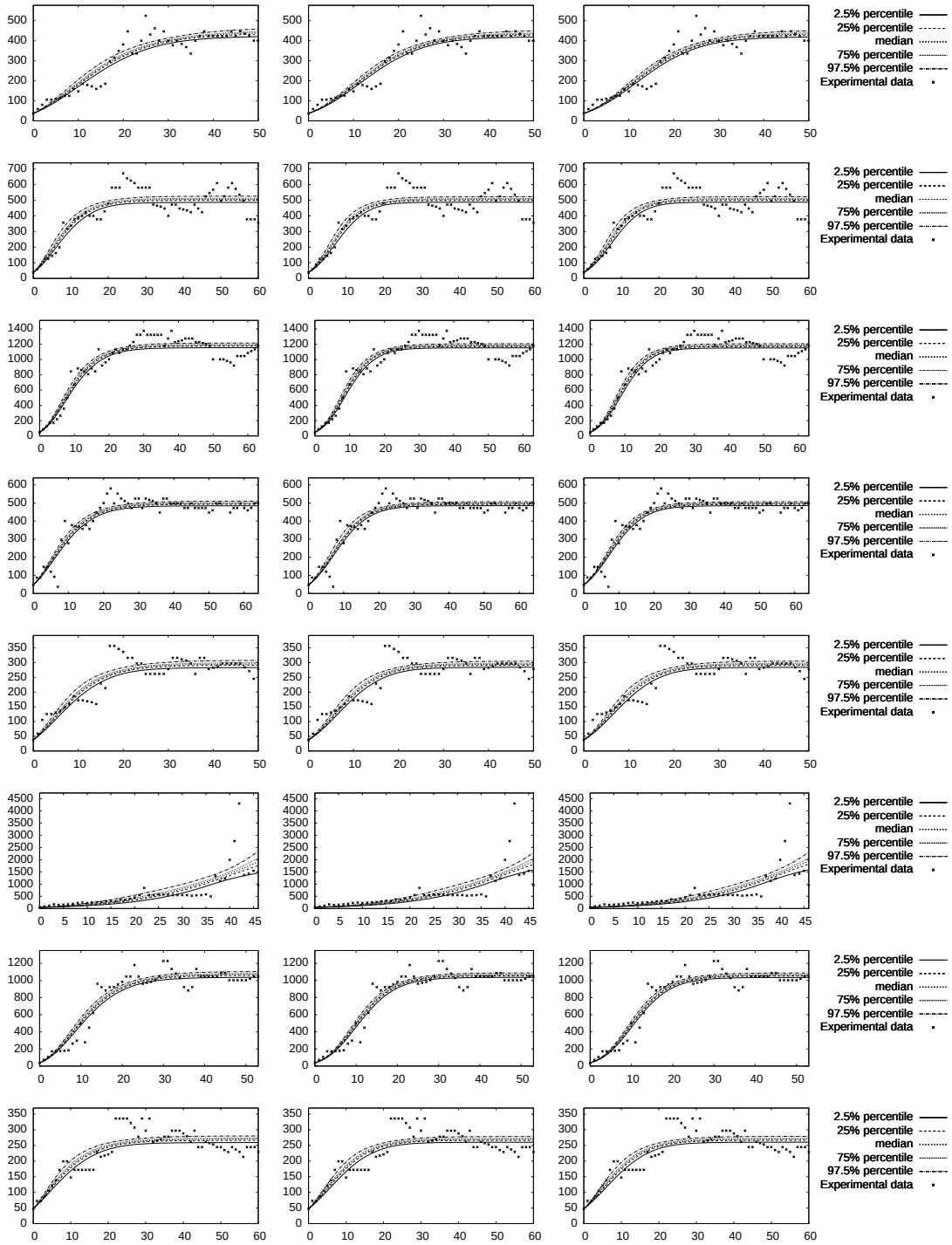

Figure S5: Predictive posteriors for the data coming from 9l cell line experiments – all plots show the spheroid volume (10^{-3}mm^3) vs. time (days). The plots on the left depict predictive posteriors produced using the traditional Gompertz model, the central ones are produced using the new model with constant λ, while the right ones are produced using the new model with variable λ. Black crosses correspond to the original data measurements, while lines show percentiles of model predictions. Each row corresponds to a separate observed spheroid (here, spheroids 1 to 8).

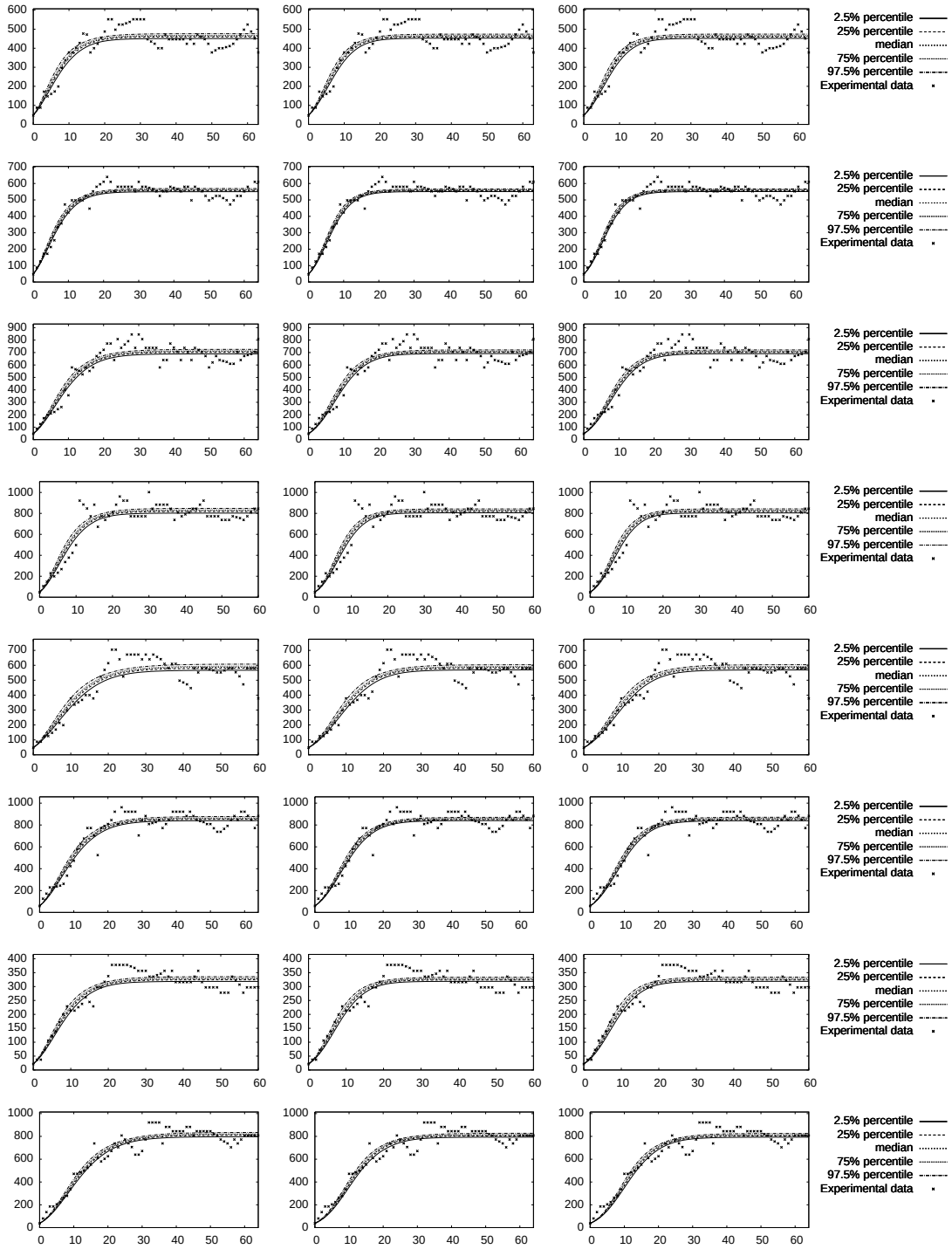

Figure S5: (continued) Predictive posteriors for the data coming from 9l cell line experiments – all plots show the spheroid volume (10^{-3}mm^3) vs. time (days). The plots on the left depict predictive posteriors produced using the traditional Gompertz model, the central ones are produced using the new model with constant λ, while the right ones are produced using the new model with variable λ. Black crosses correspond to the original data measurements, while lines show percentiles of model predictions. Each row corresponds to a separate observed spheroid (here, spheroids 9 to 16).

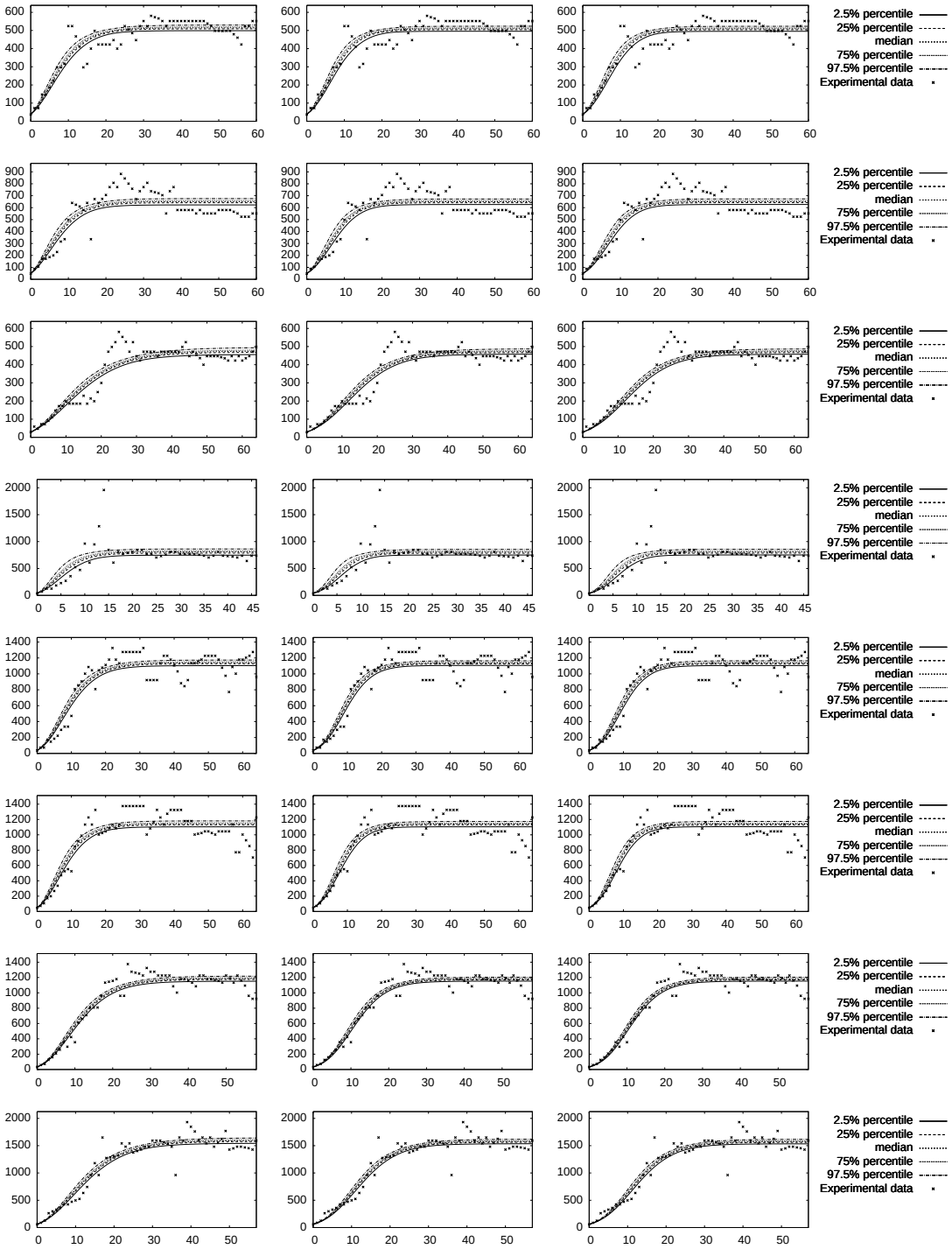

Figure S5: (continued) Predictive posteriors for the data coming from 9l cell line experiments – all plots show the spheroid volume (10^{-3}mm^3) vs. time (days). The plots on the left depict predictive posteriors produced using the traditional Gompertz model, the central ones are produced using the new model with constant λ, while the right ones are produced using the new model with variable λ. Black crosses correspond to the original data measurements, while lines show percentiles of model predictions. Each row corresponds to a separate observed spheroid (here, spheroids 17 to 24).

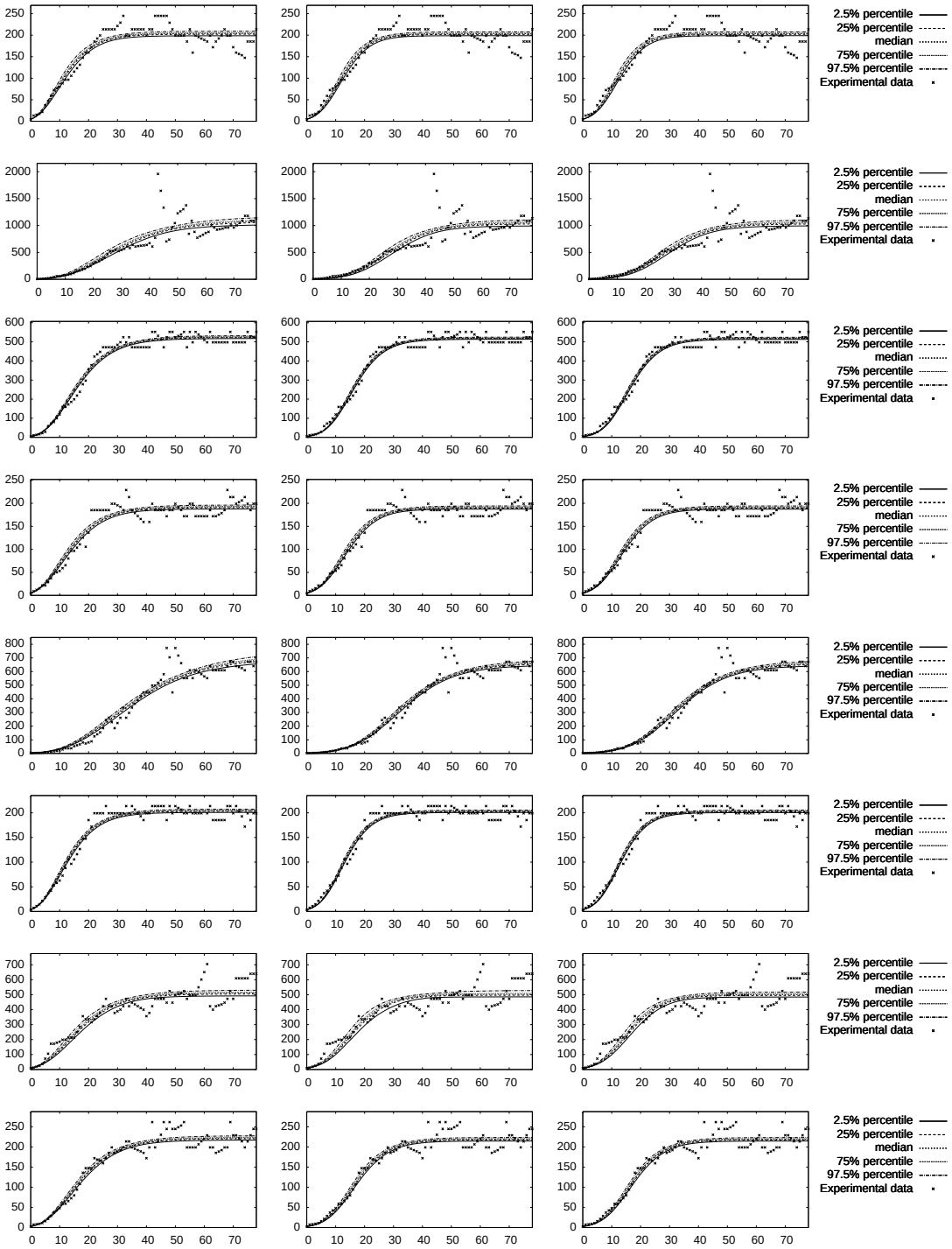

Figure S5: (continued) Predictive posteriors for the data coming from 9l cell line experiments – all plots show the spheroid volume (10^{-3}mm^3) vs. time (days). The plots on the left depict predictive posteriors produced using the traditional Gompertz model, the central ones are produced using the new model with constant λ, while the right ones are produced using the new model with variable λ. Black crosses correspond to the original data measurements, while lines show percentiles of model predictions. Each row corresponds to a separate observed spheroid (here, spheroids 25 to 32).

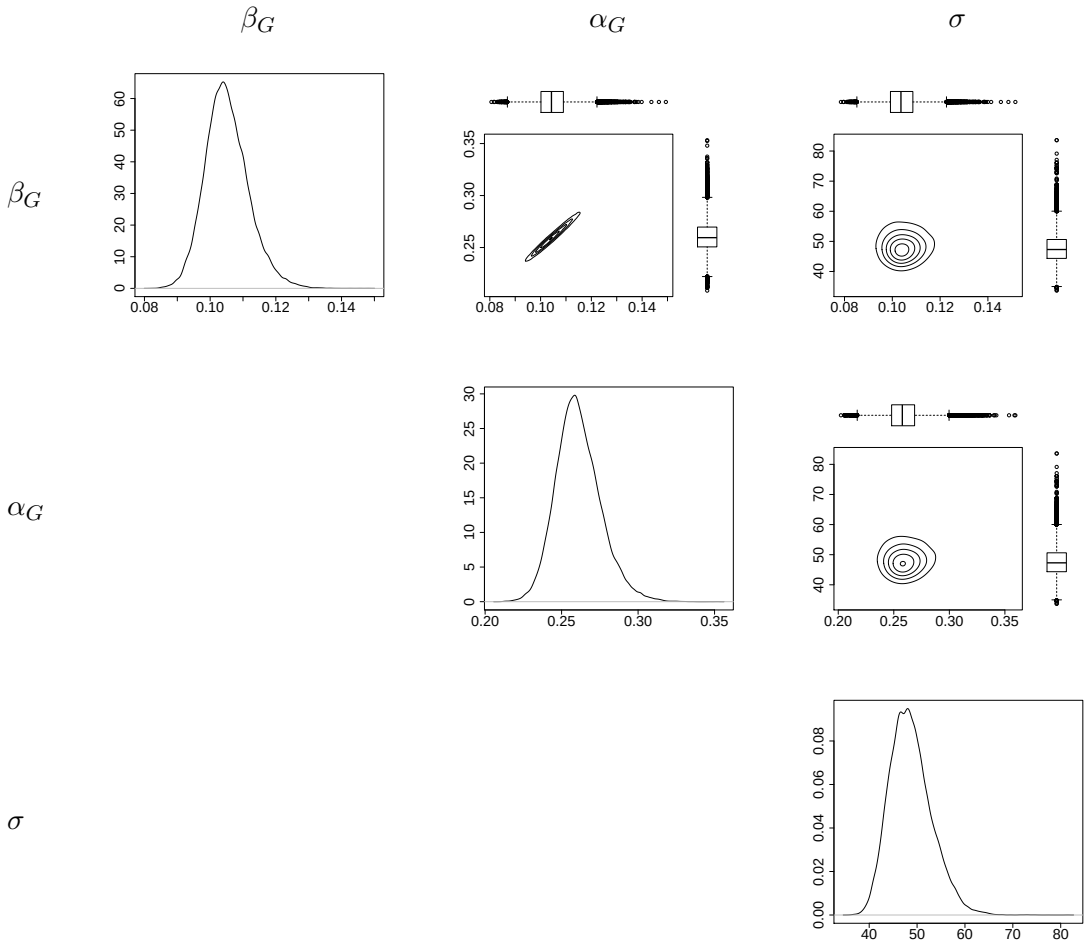

Figure S6: Marginals of the parameter posterior for the Gompertz model using the data set from the first spheroid of the 9l cell line. The α_G and β_G parameters are in days^{-1} and σ is in 10^{-3}mm^3.

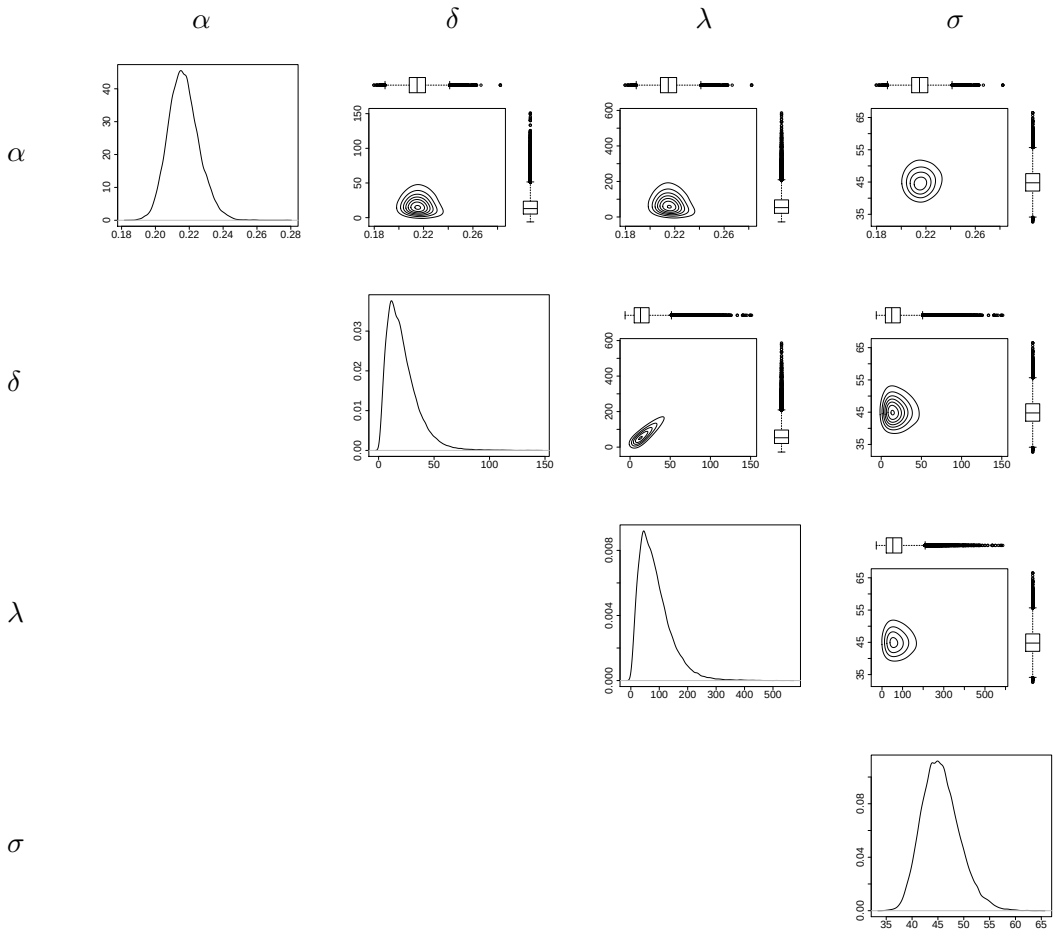

Figure S7: Marginals of the parameter posterior for the new model with constant λ, using the data set from the first spheroid of the 9l cell line. The α and δ parameters are in days^{-1}, λ is in μm and σ is in 10^{-3}mm^3.

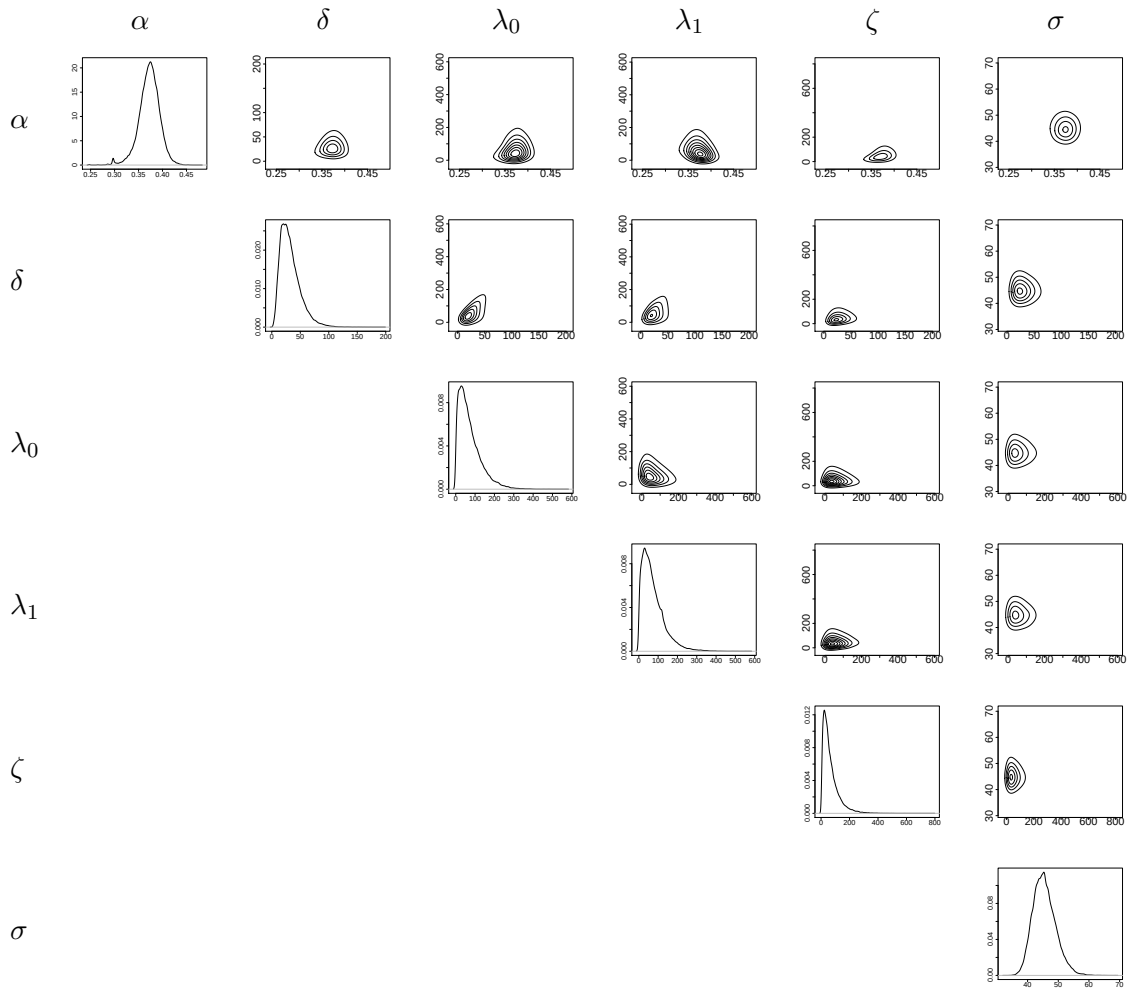

Figure S8: Marginals of the parameter posterior for the new model with variable λ, using the data set from the first spheroid of the 9l cell line. The α and δ parameters are in days^{-1}, λ_0, λ_1 and ζ are in μm and σ is in 10^{-3}mm^3.

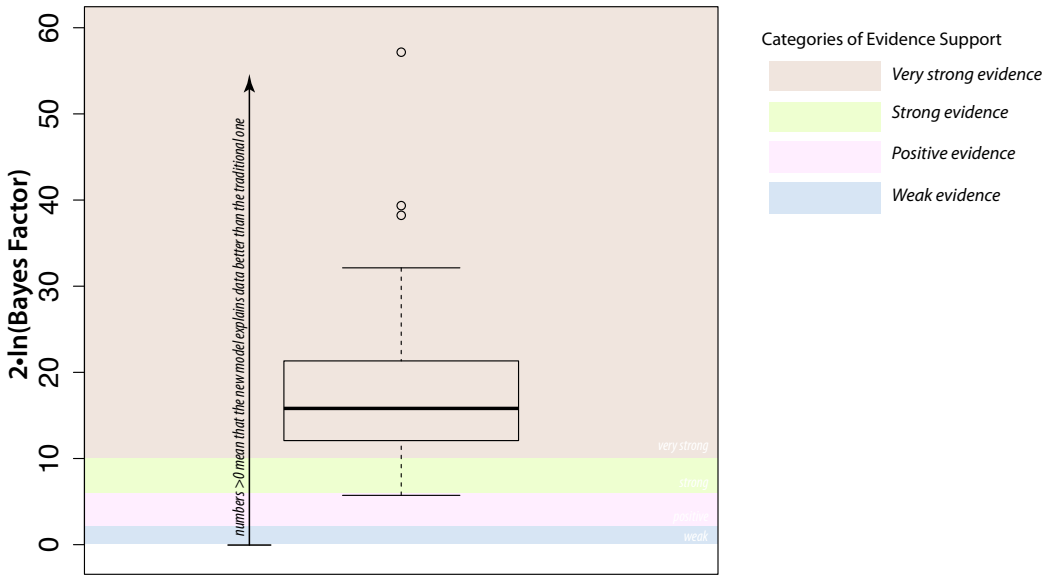

Figure S9: Bayes factors computed using the data from 32 spheroids of the 9l cell line. The box plot demonstrates that every single Bayes factor prefers the new model with constant λ over the traditional Gompertz model. The categories of evidence support defined in Table S2 are plotted against our results.

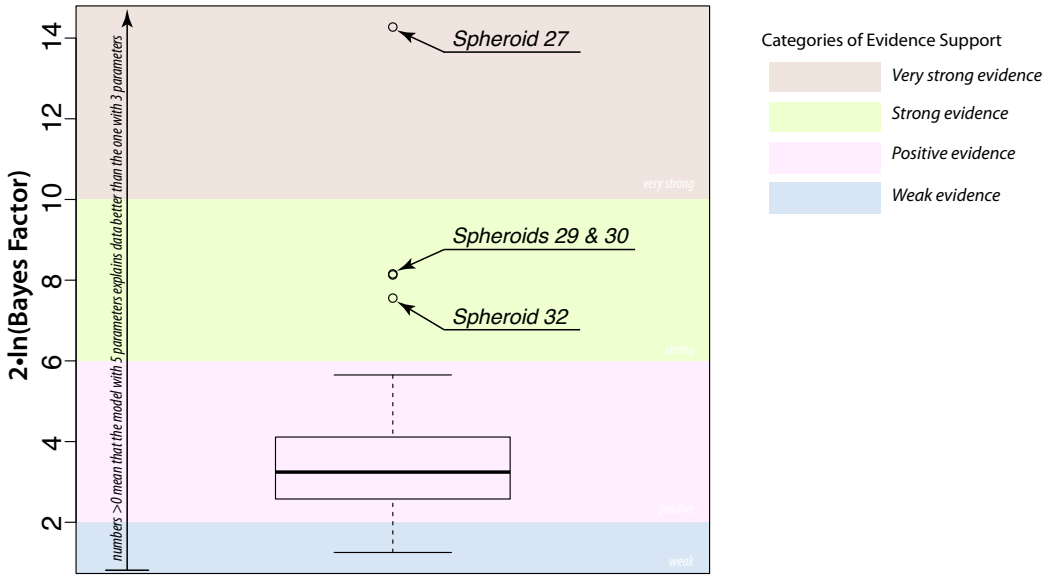

Figure S10: Bayes factors computed using the data from 32 spheroids of the 9l cell line. The box plot demonstrates that the new model with variable λ is slightly better supported by the data than the one with constant λ. The categories of evidence support defined in Table S2 are plotted against our results.

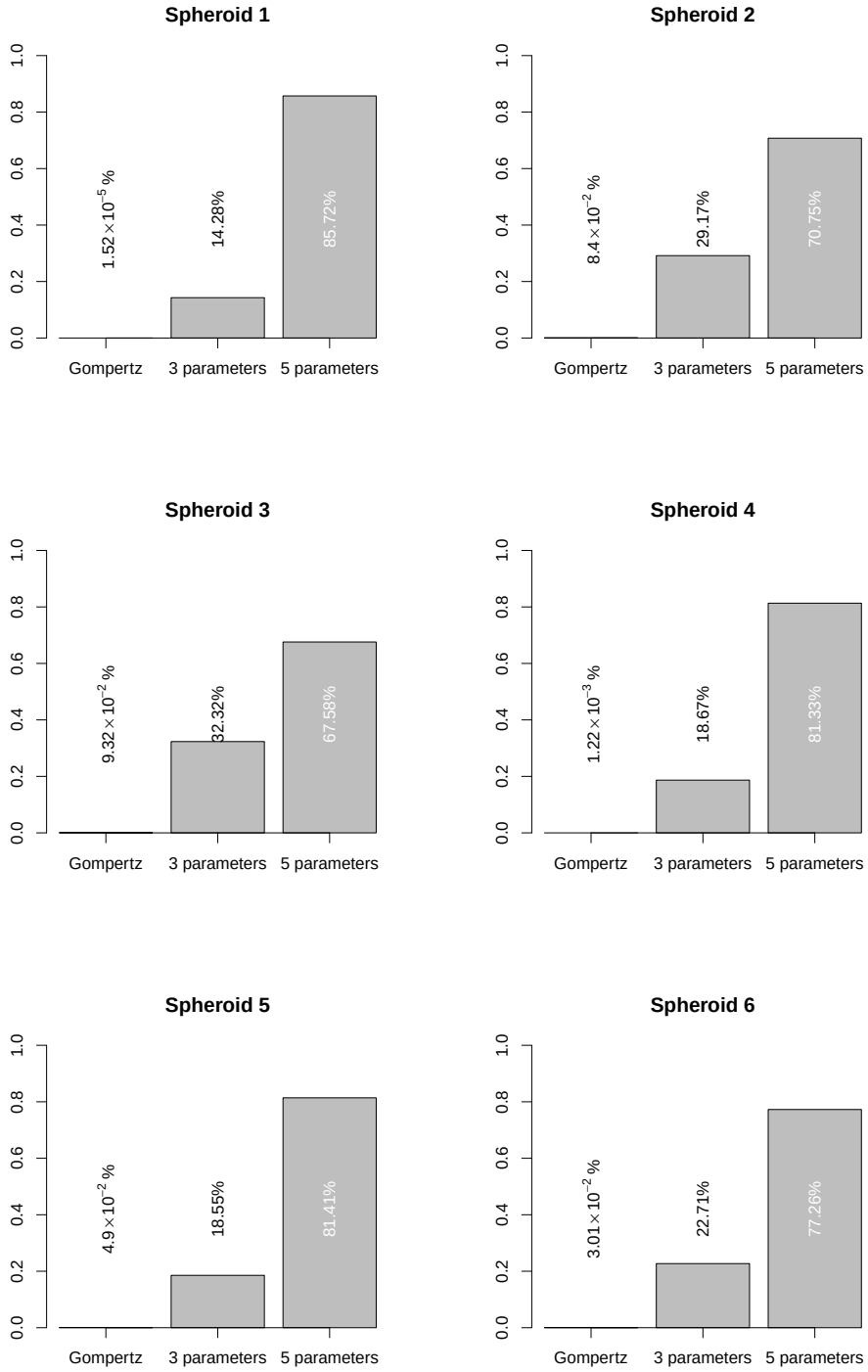

Figure S11: A posteriori odds of the alternative models given data from 9l cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

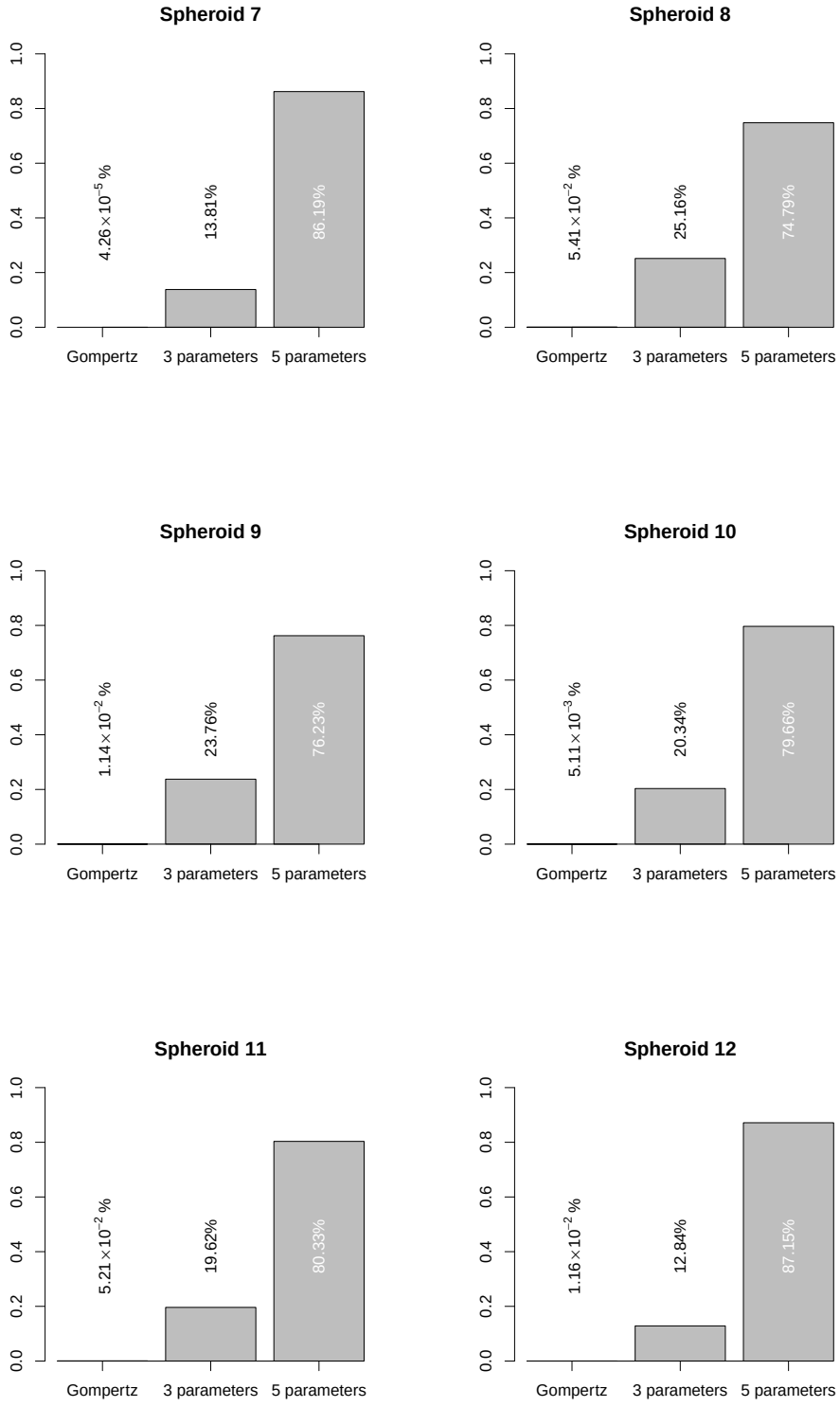

Figure S11: (continued) A posteriori odds of the alternative models given data from 9l cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

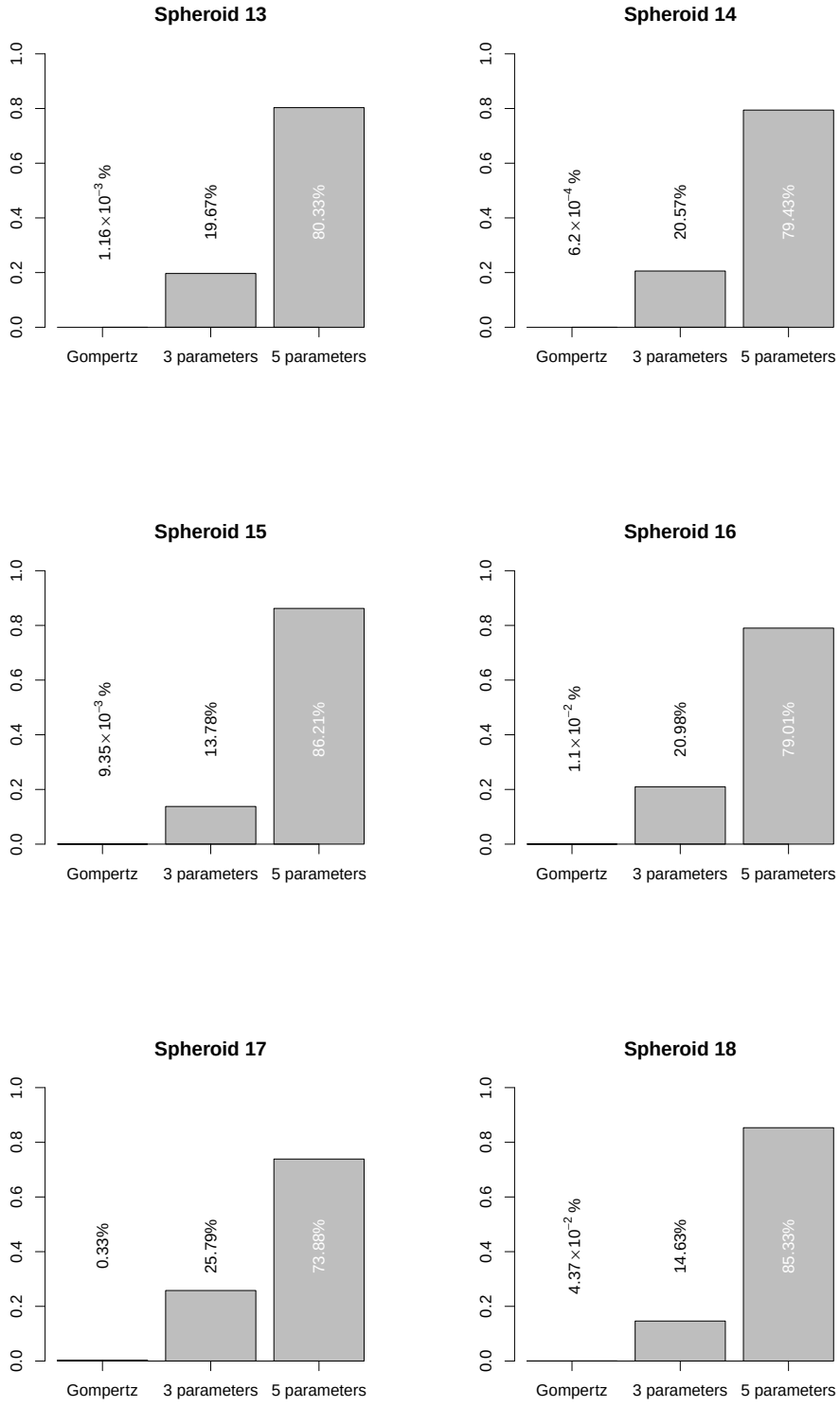

Figure S11: (continued) A posteriori odds of the alternative models given data from 9l cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

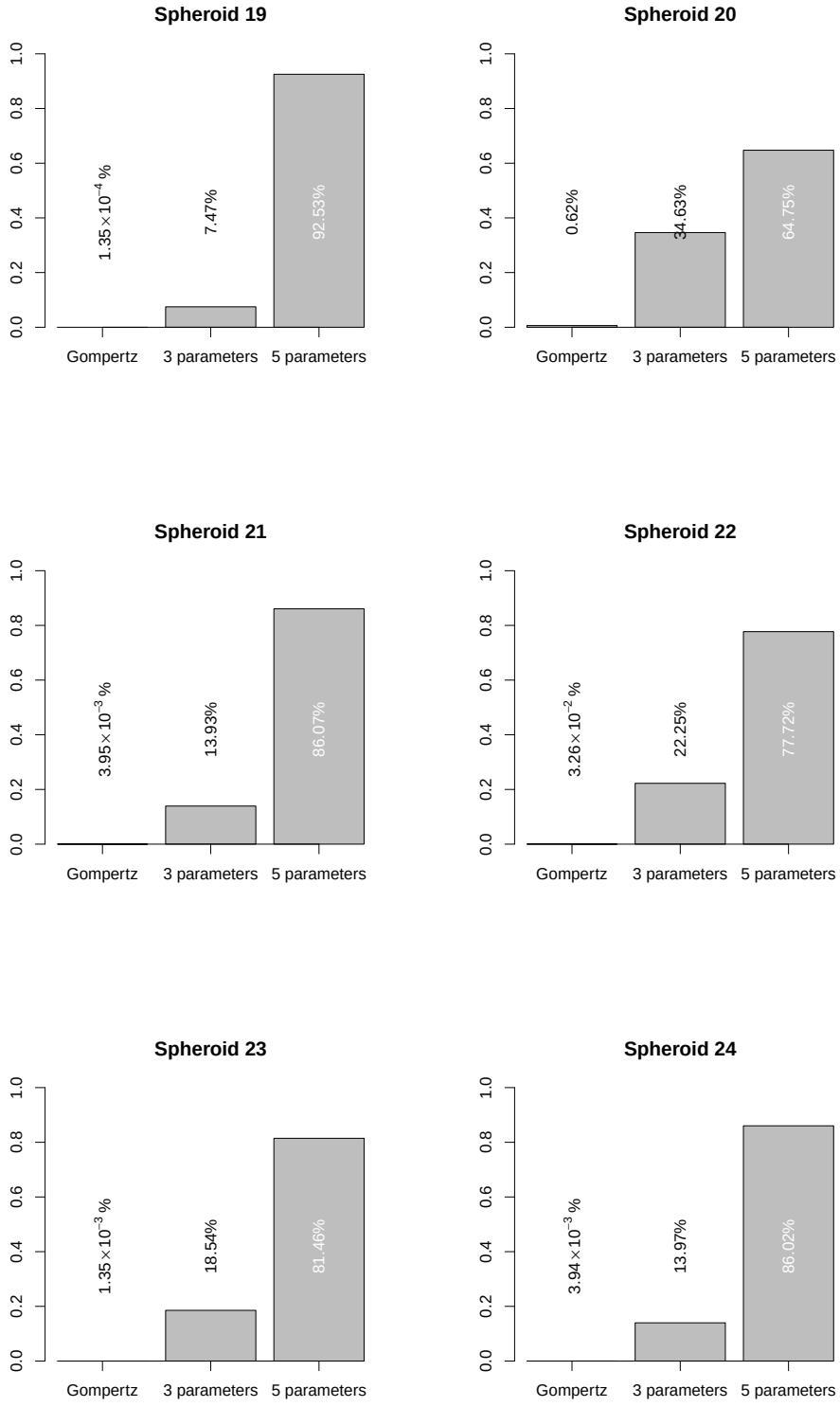

Figure S11: (continued) A posteriori odds of the alternative models given data from 9l cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

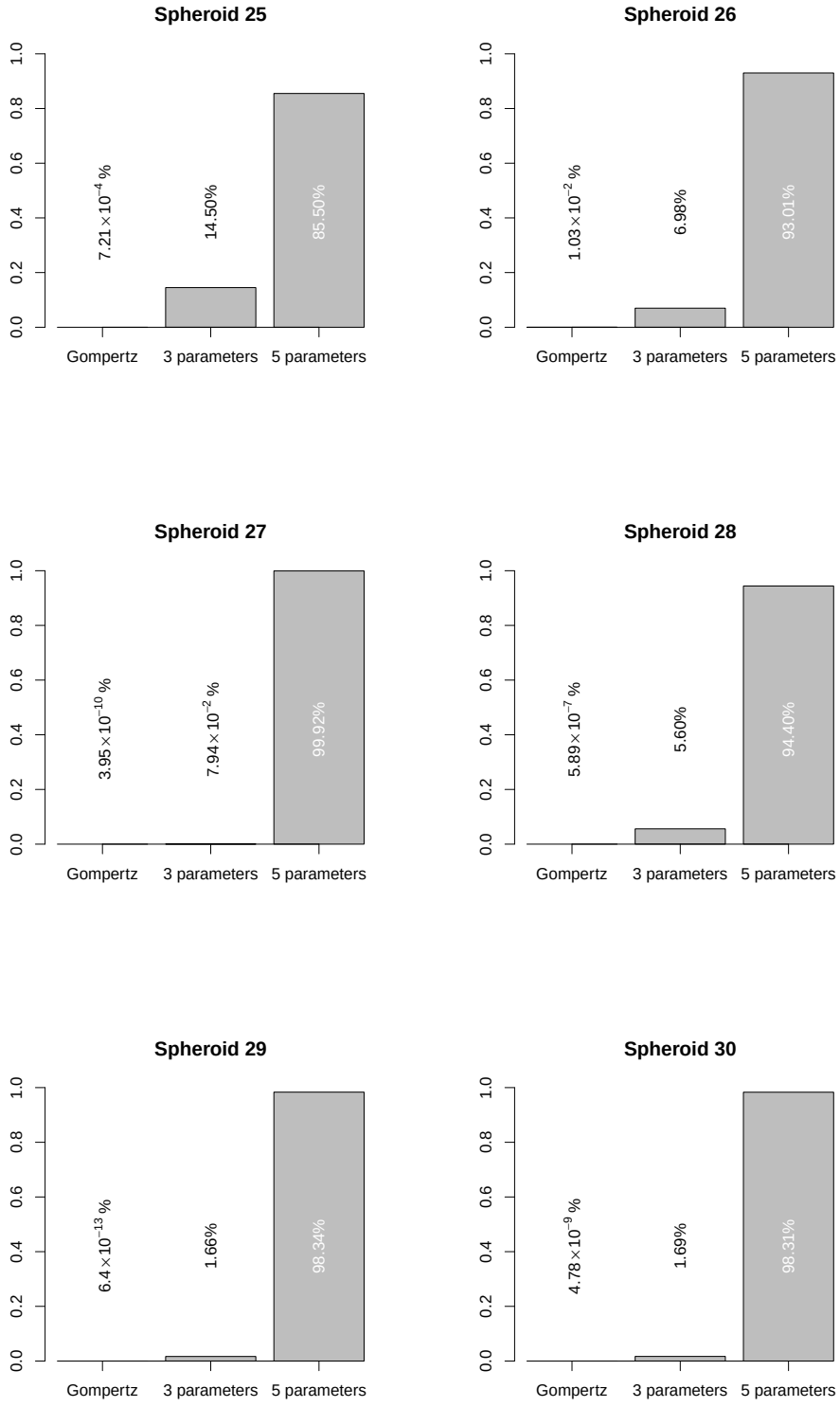

Figure S11: (continued) A posteriori odds of the alternative models given data from 9l cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

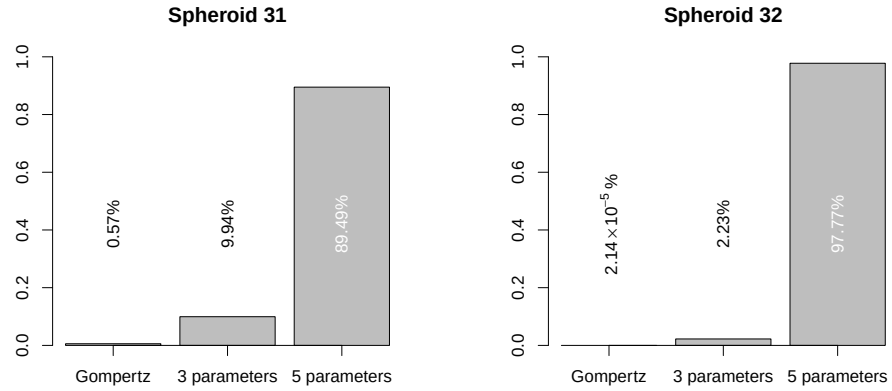

Figure S11: (continued) A posteriori odds of the alternative models given data from 9l cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

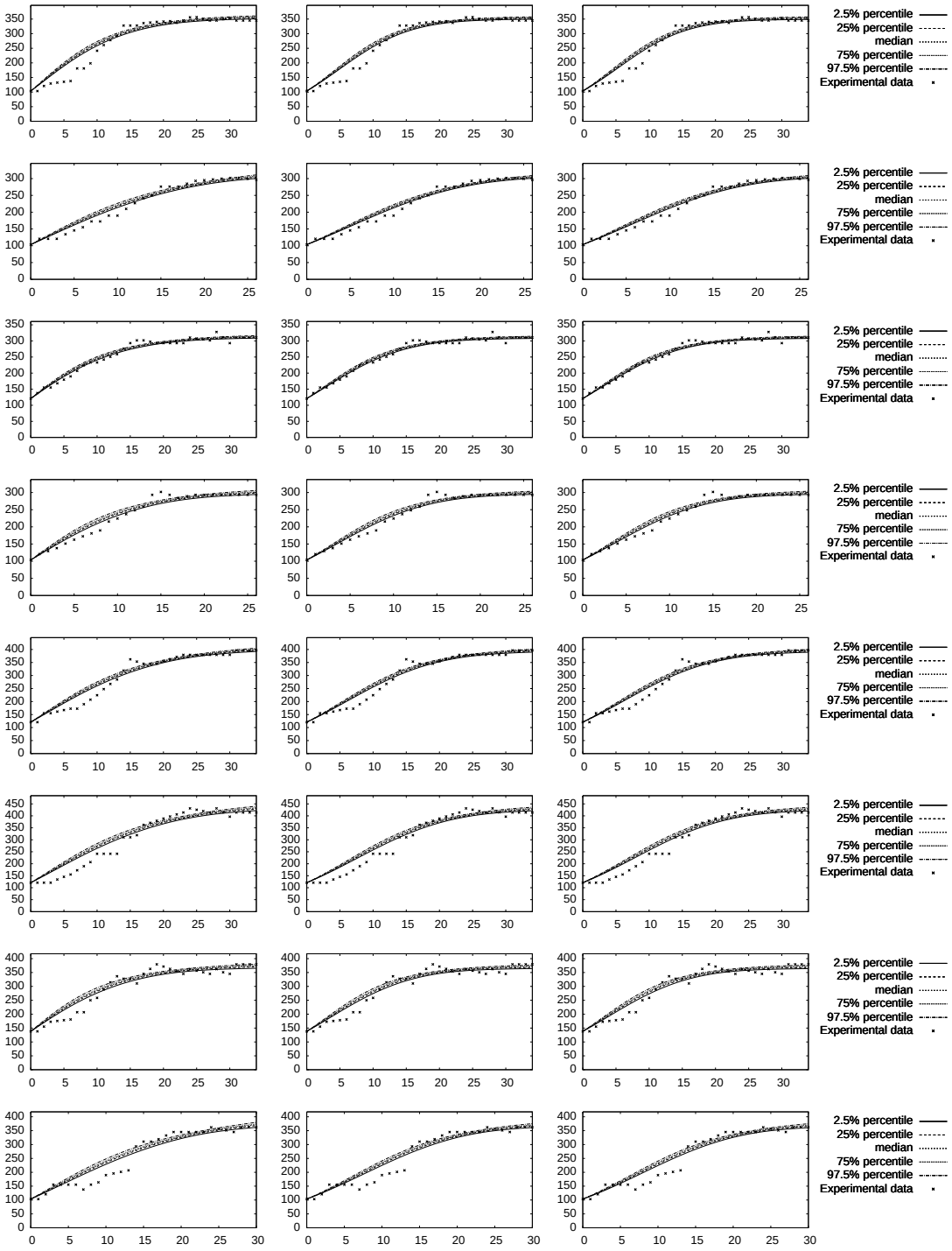

Figure S12: Predictive posteriors for the data from U118 cell line experiments – all plots show the spheroid radius (μm) vs. time (days). The plots on the left depict predictive posteriors produced using the traditional Gompertz model, the central ones are produced using the new model with constant λ, while the right ones are produced using the new model with variable λ. Black crosses correspond to the original data measurements, while lines show percentiles of model predictions. Each row corresponds to a separate observed spheroid (here, spheroids 1 to 8).

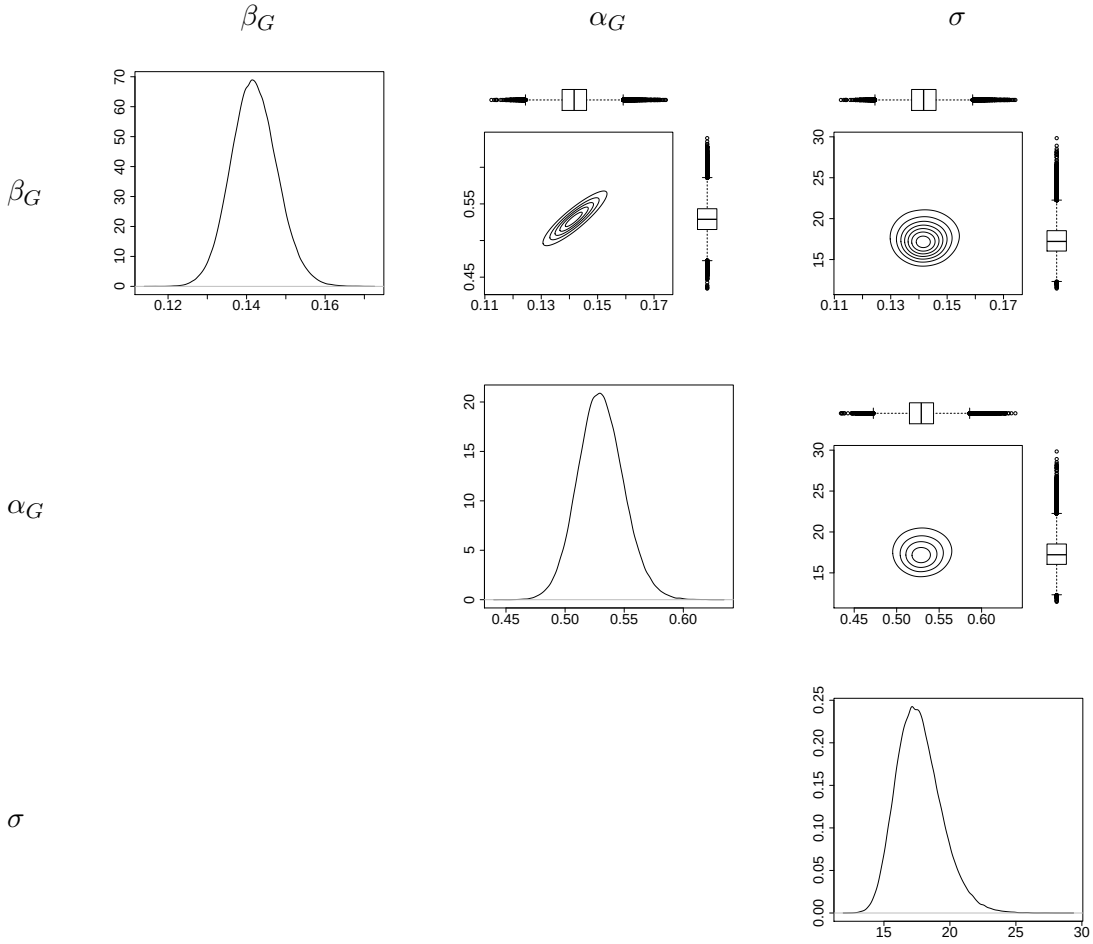

Figure S13: Marginals of the parameter posterior for the Gompertz model using the data set from the first spheroid of the U118 cell line. The α_G and β_G parameters are in days^{-1} and σ is in μm.

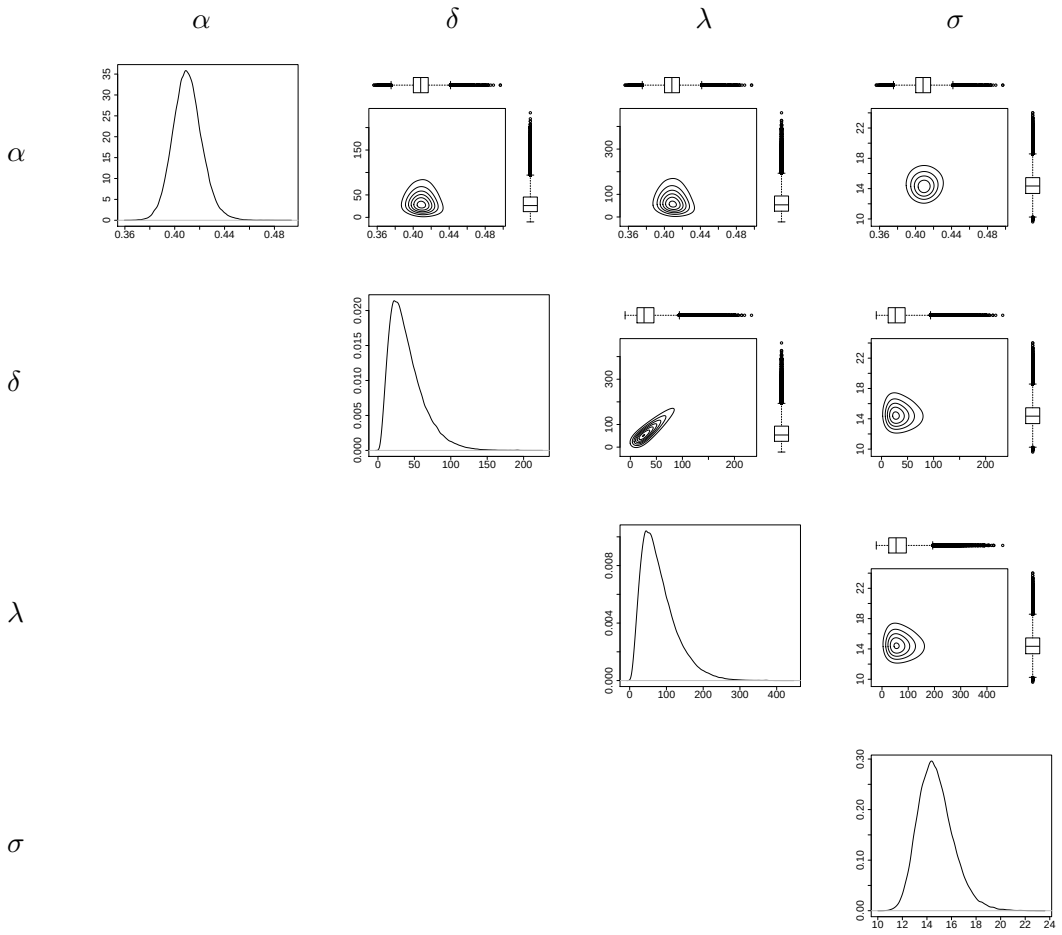

Figure S14: Marginals of the parameter posterior for the new model with constant λ using the data set from the first spheroid of the U118 cell line. The α and δ parameters are in days^{-1}, λ and σ are in μm.

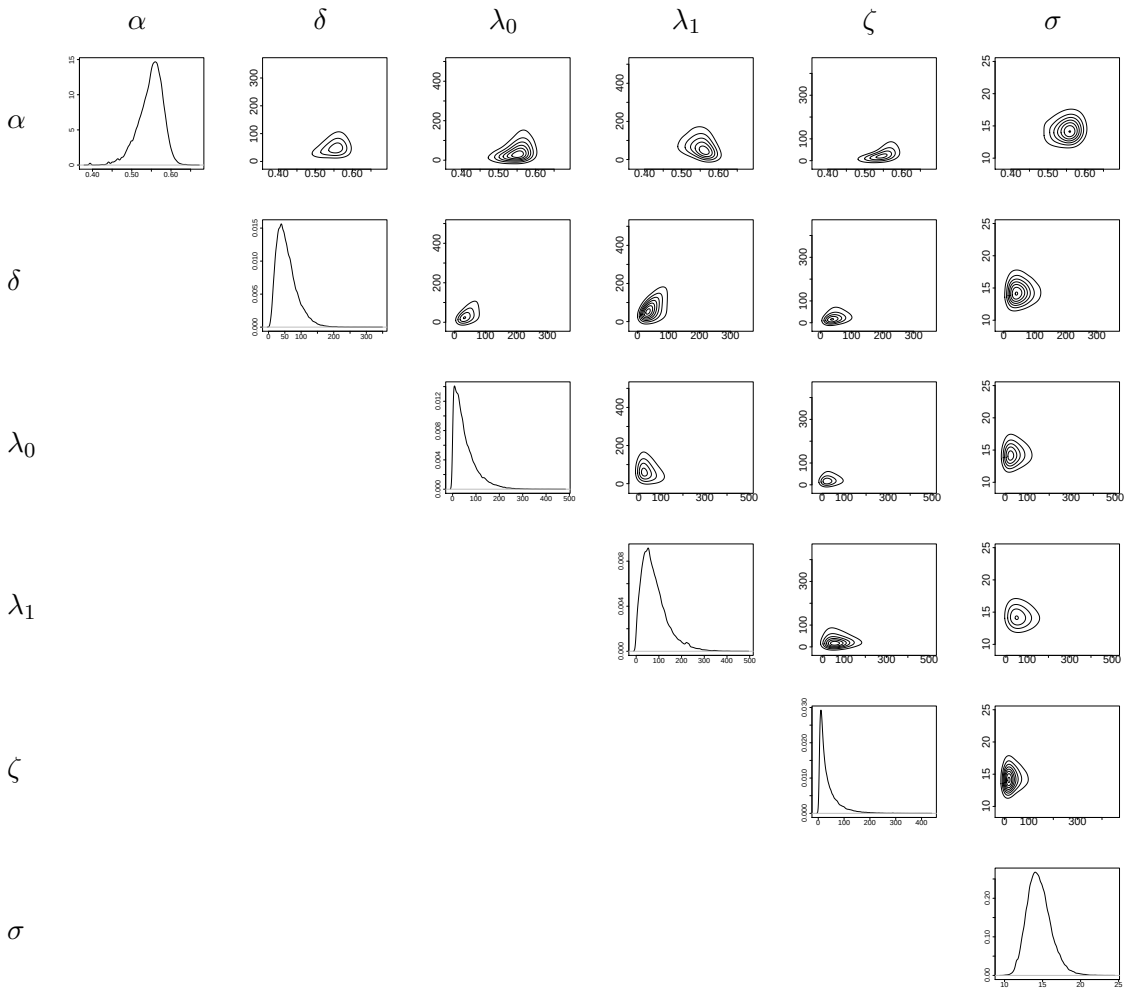

Figure S15: Marginals of the parameter posterior for the new model with variable λ using the data set from the first spheroid of the U118 cell line. The α and δ parameters are in days^{-1}, λ_0, λ_1, ζ and σ are in μm.

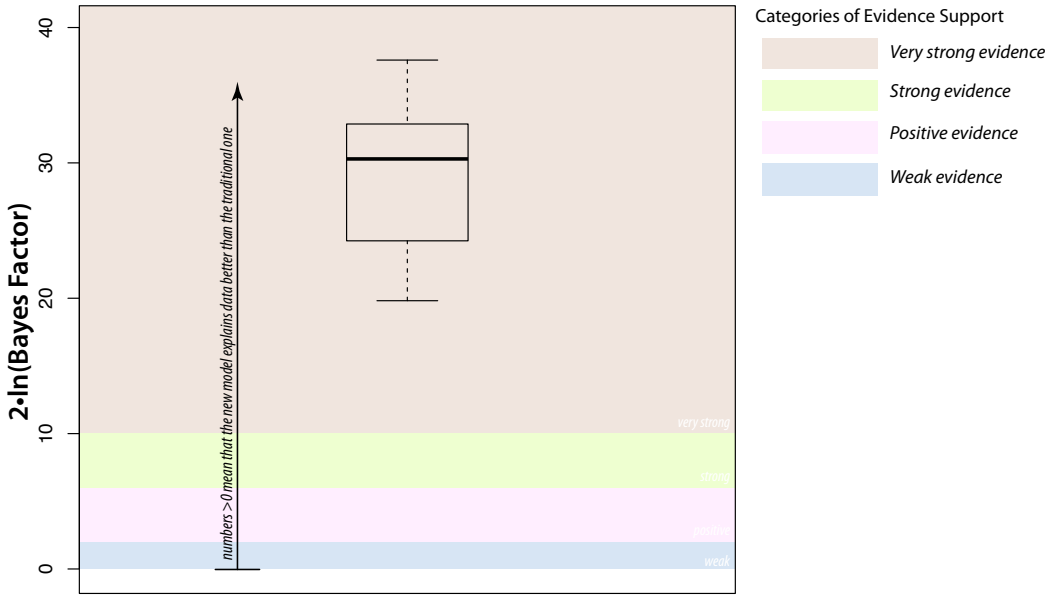

Figure S16: Bayes factors computed using the data from 8 spheroids of the U118 cell line. The box plot demonstrates that every single Bayes factor very strongly prefers the new model over the traditional Gompertz model. The categories of evidence support defined in Table S2 are plotted against our results.

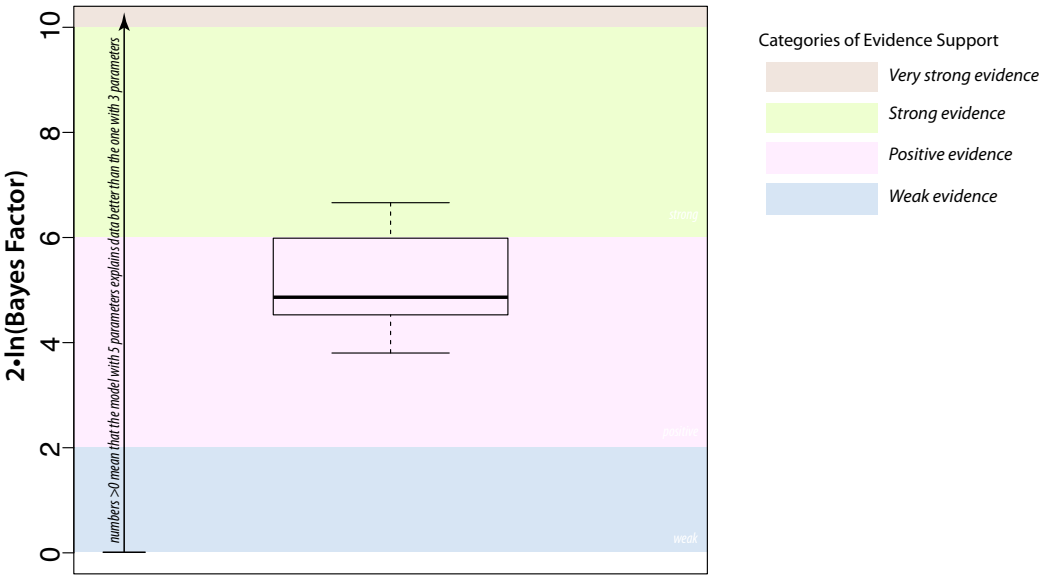

Figure S17: Bayes factors computed using the data from 8 spheroids of the U118 cell line. The box plot demonstrates that the new model with variable λ is slightly better supported by the data than the one with constant λ. The categories of evidence support defined in Table S2 are plotted against our results.

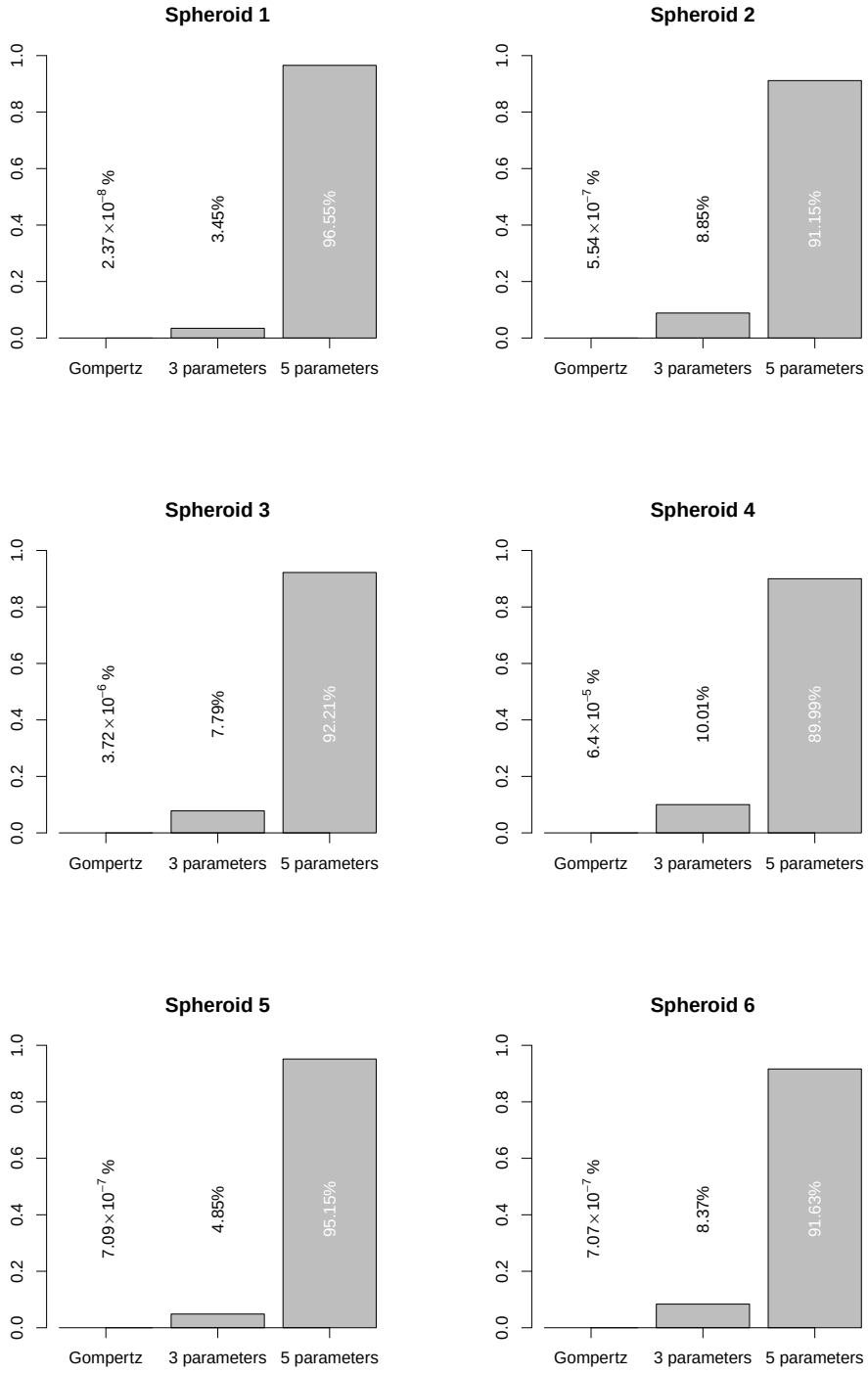

Figure S18: A posteriori odds of the alternative models given data from U118 cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

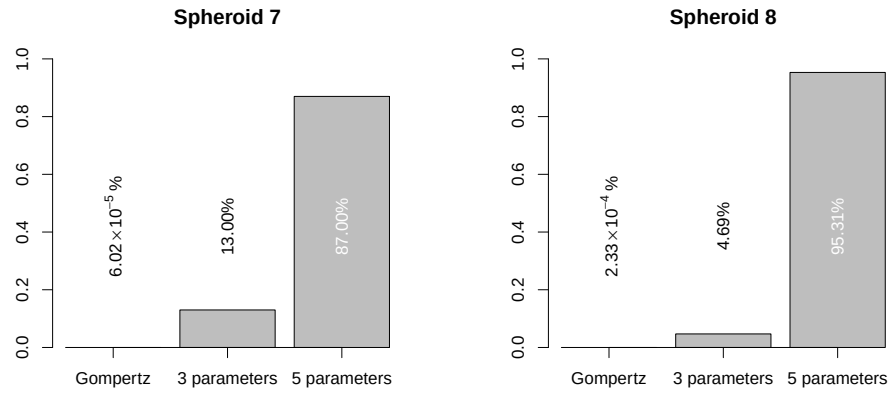

Figure S18: (continued) A posteriori odds of the alternative models given data from U118 cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

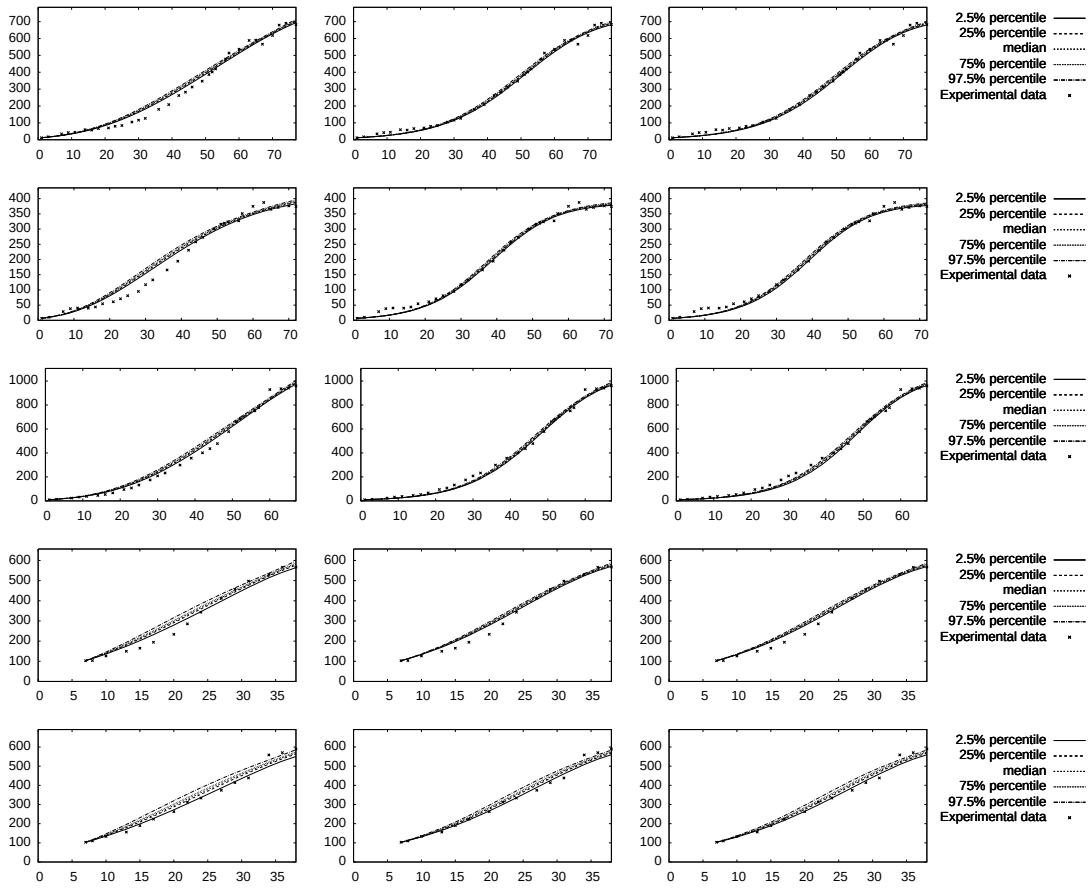

Figure S19: Predictive posteriors for the data from MCF7 cell line experiments – all plots show the spheroid radius (μm) vs. time (days). The plots on the left depict predictive posteriors produced using the traditional Gompertz model, the central ones are produced using the new model with constant λ, while the right ones are produced using the new model with variable λ. Black crosses correspond to the original data measurements, while lines show percentiles of model predictions. Each row corresponds to a separate spheroid observed (spheroids 1 to 5).

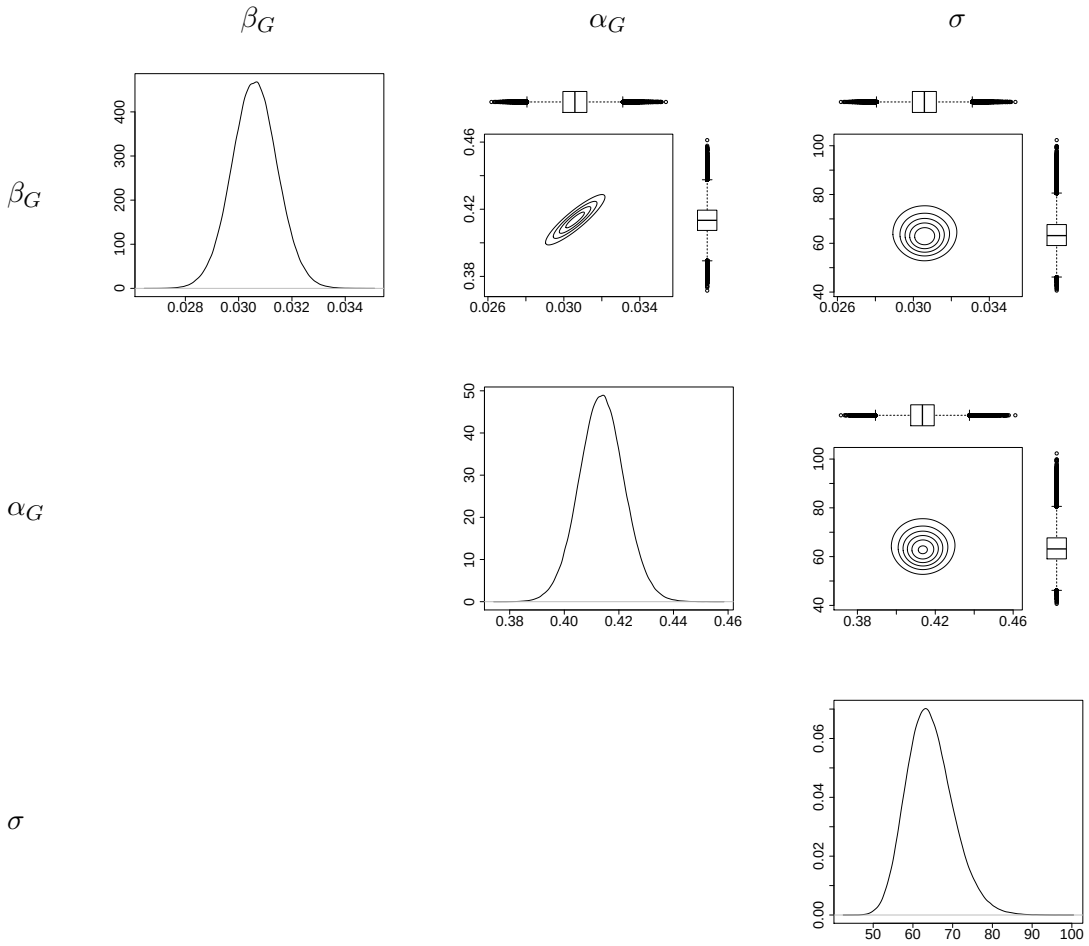

Figure S20: Marginals of the parameter posterior for the Gompertz model using the data set from the first spheroid of the MCF cell line. The α_G and β_G parameters are in days^{-1} and σ is in μm.

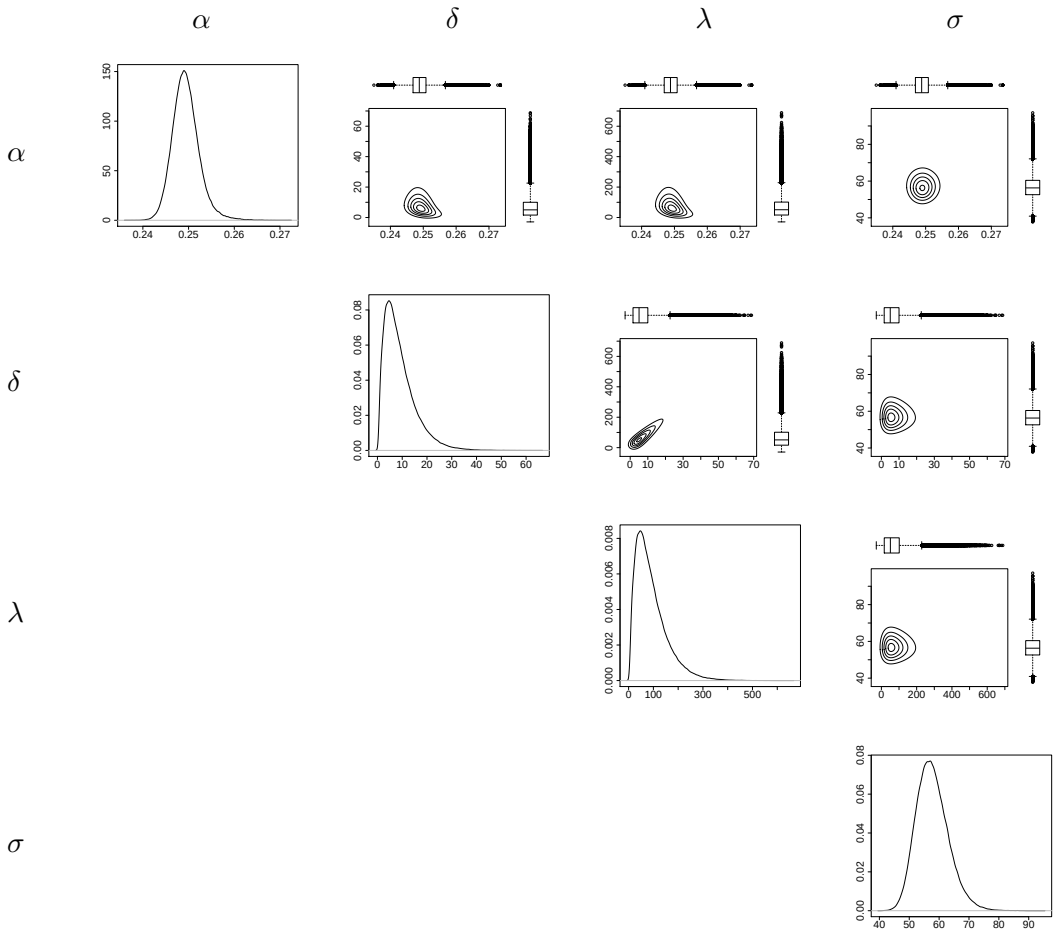

Figure S21: Marginals of the parameter posterior for the new model using the data set from the first spheroid of the MCF7 cell line. The α and δ parameters are in days^{-1}, λ and σ are in μm.

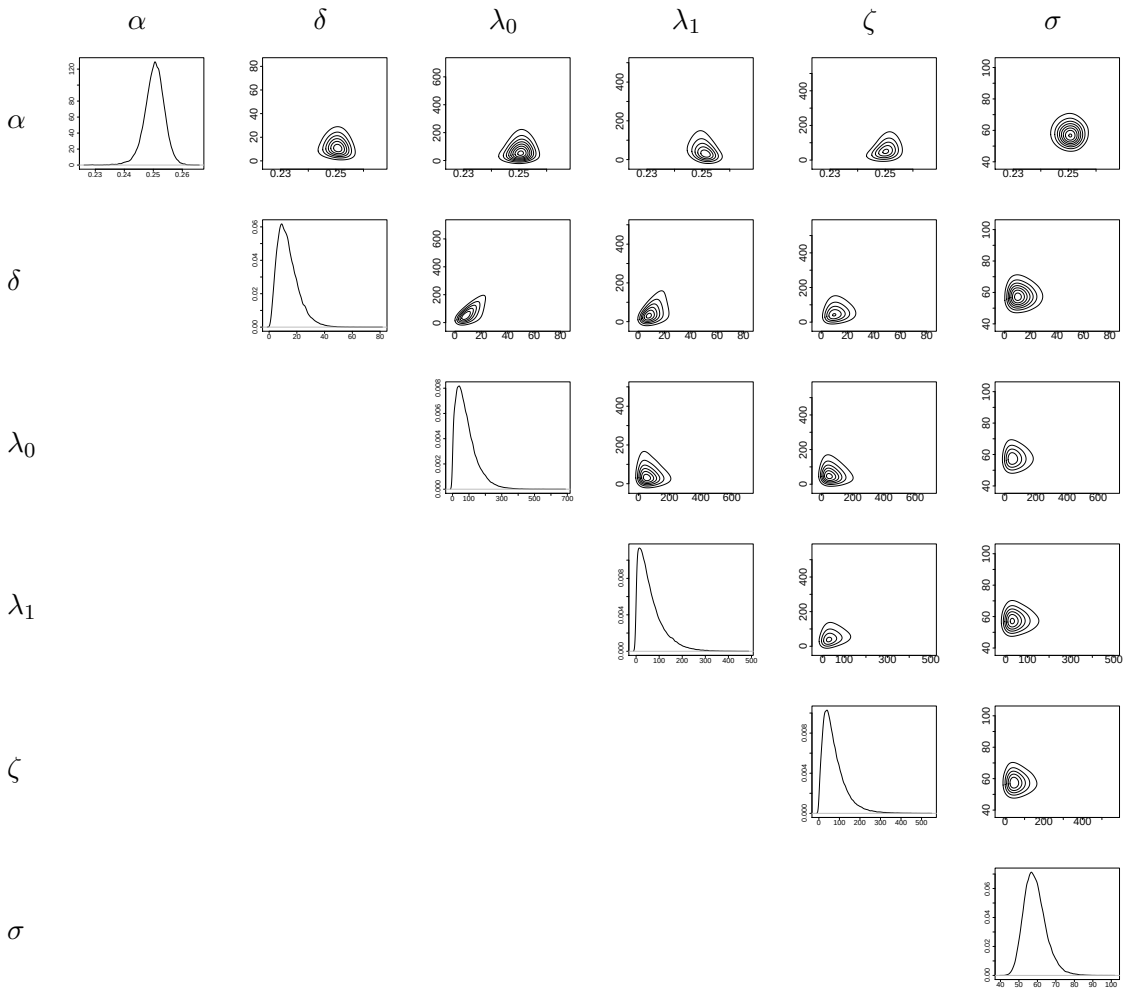

Figure S22: Marginals of the parameter posterior for the new model with variable λ using the data set from the first spheroid of the MCF7 cell line. The α and δ parameters are in days^{-1}, λ_0, λ_1, ζ and σ are in μm.

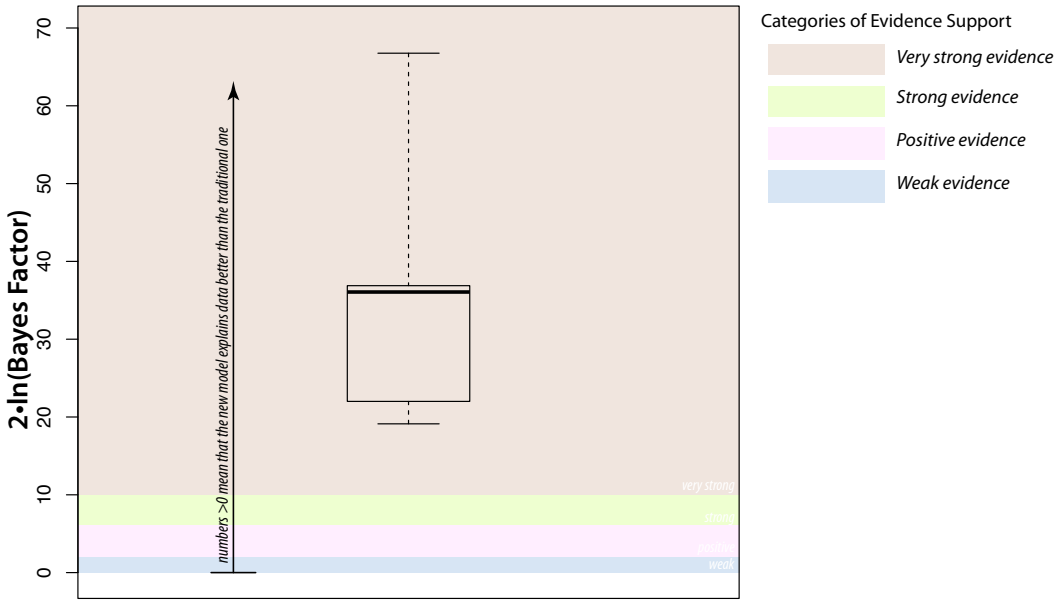

Figure S23: Bayes factors computed using the data from 5 spheroids of the MCF7 cell line. The box plot demonstrates that every single Bayes factor very strongly prefers the new model over the traditional Gompertz model. The categories of evidence support defined in Table S2 are plotted against our results.

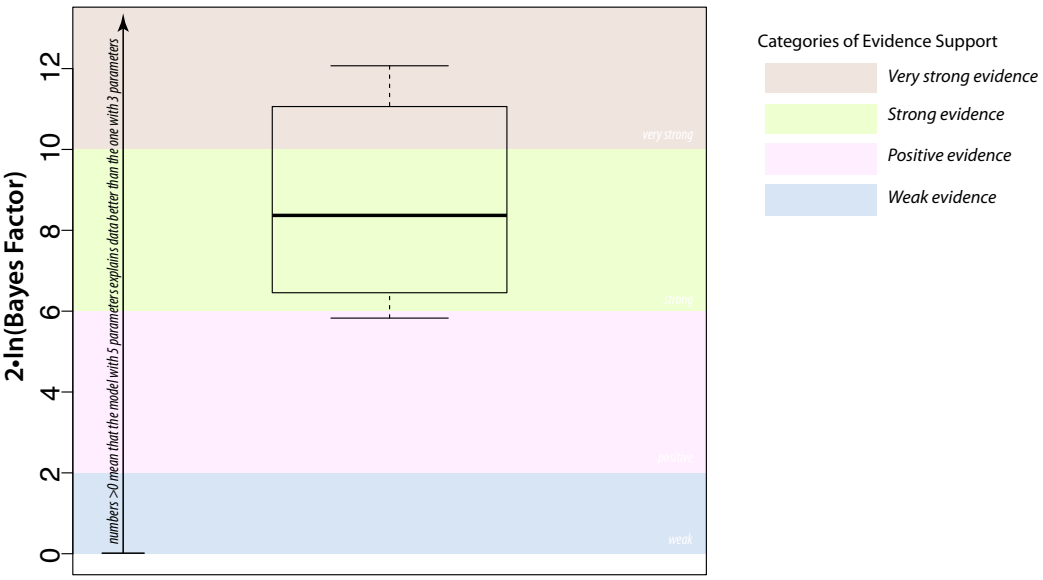

Figure S24: Bayes factors computed using the data from 8 spheroids of the MCF7 cell line. The box plot demonstrates that the new model with variable λ is better supported by the data than the one with constant λ. The categories of evidence support defined in Table S2 are plotted against our results.

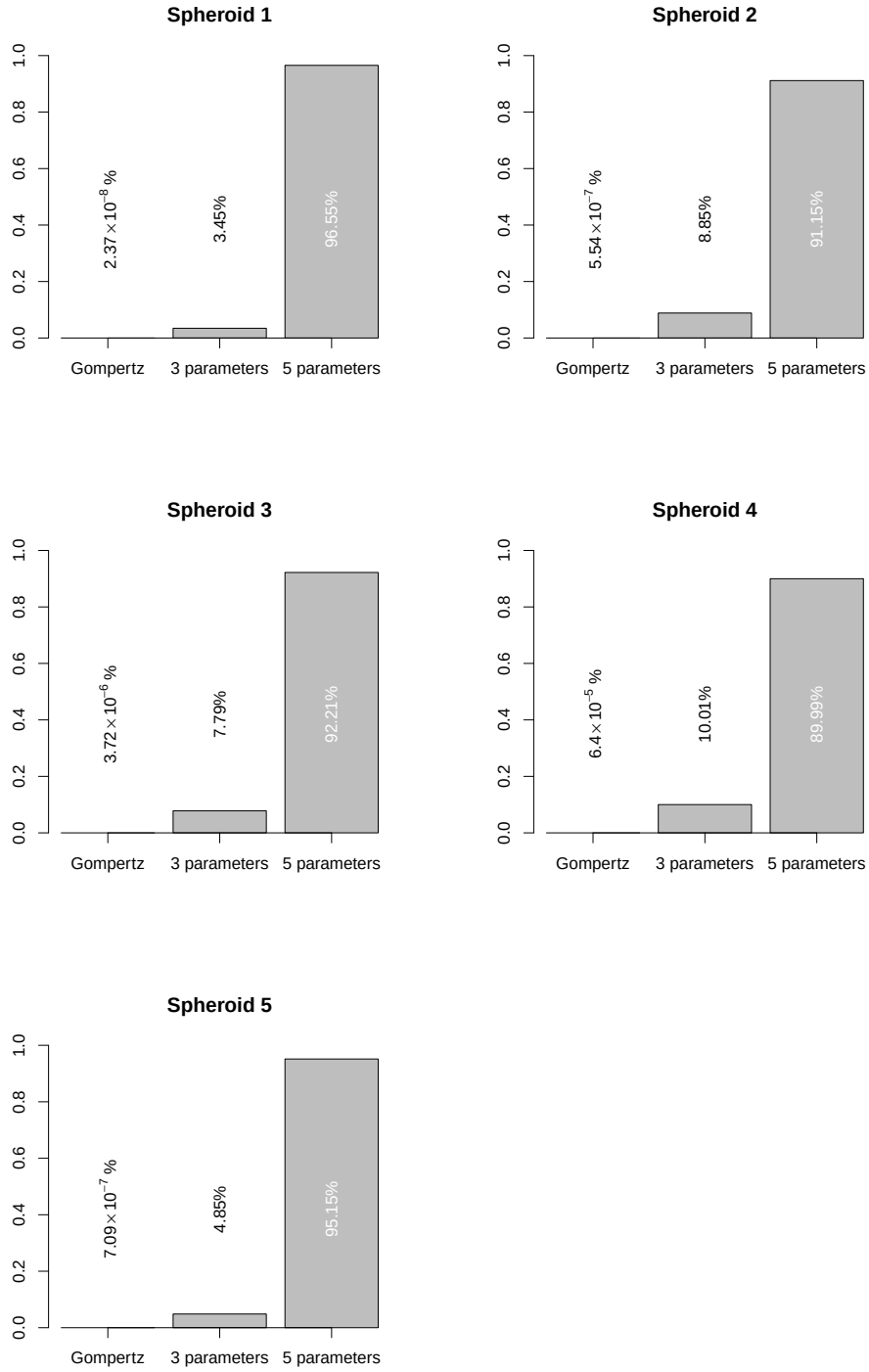

Figure S25: A posteriori odds of the alternative models given data from MCF7 cell line demonstrate that new models are significantly preferred to the traditional Gompertz model, while the evidence is not very high to decisively prefer the new model with variable λ over the one with constant λ.

References

Arkinson, A. C. (1978). Posterior probabilities for choosing a regression model. *Biometrika 65*, 39–48.

Bernardo, J. M. and A. F. M. Smith (1994). *Bayesian Theory*. Wiley, Chichester.

Chib, S. (1995). Marginal likelihood from the Gibbs output. *J. Amer. Statist. Assoc. 90*(432), 1313–1321.

Cox, D. R. and D. V. Hinkley (1974). *Theoretical Statistics*. London: Chapman and Hall.

Del Moral, P., A. Doucet, and A. Jasra (2006). Sequential Monte Carlo samplers. *J. R. Statist. Soc. B 68*(Part 3), 411–436.

Freedman, D. A. (1983). A note on screening regression equations. *The American Statistician 37*, 152–155.

Friel, N. and A. N. Pettitt (2006). Marginal likelihood estimation via power posteriors. Technical report, Department of Statistics, University of Glasgow.

Gelman, A. (1998). Simulating normalising constants: From importance sampling to bridge sampling to path sampling. *Stat. Sci. 13*, 163–185.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin (1995). *Bayesian Data Analysis*. Chapman & Hall.

Gilks, W., S. Richardson, and D. Spiegelhalter (1995). *Markov Chain Monte Carlo in Practice*. Chapman&Hall/CRC.

Jeffreys, H. (1961). *Theory of Probability (3rd ed.)*. Oxford University Press.

Kass, R. E. and A. E. Raftery (1995). Bayes factors. *Journal of the American Statistical Association 90*(430), 773–795.

Lartillot, N. and H. Philippe (2006). Computing Bayes factors using thermodynamic integration. *Systematic Biology 55*(2), 195–207.

Neal, R. (1993). Probabilistic inference using Markov Chain Monte Carlo methods. Technical Report CRG-TR-93-1, Dept. of Computer Science, University of Toronto.

Neal, R. M. (2001). Annealed importance sampling. *Statistics and Computing 11*, 125–139.

Newton, M. A. and A. E. Raftery (1994). Approximate Bayesian inference by the weighted likelihood bootstrap. *JRSS Ser. B 3*, 3–48.

Ogata, Y. (1989). A Monte Carlo method for high dimensional integration. *Num. Math. 55*, 137–157.

Raftery, A. E. (1986). Choosing models for cross-classifications. *American Sociological Review 51*, 145–146.

Raftery, A. E., D. M. Madigan, and J. A. Hoeting (1993). Model selection and accounting for model uncertainty in linear regression models. Technical Report 262, University of Washington, Dept. of Statistics.

Robert, C. P. and G. Casella (2004). *Monte Carlo Statistical Methods (second edition)*. New York: Springer-Verlag.

Smith, A. F. M. and D. J. Spiegelhalter (1980). Bayes factors and choice criteria for linear models. *JRSS Ser. B 42*, 213–220.