Population Modeling of Tumor Growth Curves, the Reduced Gompertz Model and Prediction of the Age of a Tumor

Cristina Vaghi, Anne Rodallec, Raphaelle Fanciullino, Joseph Ciccolini, Jonathan Paul M. Mochel, Michalis Mastri, John M.L. Ebos, Clair Poignard, Sébastien Benzekry

Population modeling of tumor growth curves, the reduced Gompertz model and prediction of the age of a tumor

C. Vaghi[1,2], A. Rodallec[3], R. Fanciullino[3], J. Ciccolini[3], J. Mochel[4], M. Mastri[5], J. ML Ebos[5], C. Poignard[1,2], and S. Benzekry[1,2]

[1] MONC team, Inria Bordeaux Sud-Ouest, France
[2] Institut de Mathématiques de Bordeaux, France
[3] SMARTc, Center for Research on Cancer of Marseille, France
[4] Iowa State University, Department of Biomedical Sciences, Ames, USA
[5] Roswell Park Comprehensive Cancer Center, Buffalo, NY, USA

Abstract. Quantitative analysis of tumor growth kinetics has been widely carried out using mathematical models. In the majority of cases, individual or average data were fitted.

Here, we analyzed three classical models (exponential, logistic and Gompertz within the statistical framework of nonlinear mixed-effects modelling, which allowed us to account for inter-animal variability within a population group. We used *in vivo* data of subcutaneously implanted Lewis Lung carcinoma cells. While the exponential and logistic models failed to accurately fit the data, the Gompertz model provided a superior descriptive power. Moreover, we observed a strong correlation between the Gompertz parameters. Combining this observation with rigorous population parameter estimation motivated a simplification of the standard Gompertz model in a reduced Gompertz model, with only one individual parameter. Using Bayesian inference, we further applied the population methodology to predict the individual initiation times of the tumors from only three measurements. Thanks to its simplicity, the reduced Gompertz model exhibited superior predictive power.

The method that we propose here remains to be extended to clinical data, but these results are promising for the personalized estimation of the tumor age given limited data at diagnosis.

Keywords: Tumor growth kinetics, Gompertz model, Mixed-effects modeling, Bayesian estimation

1 Introduction

Tumor growth kinetics have been studied since several decades both clinically [8] and experimentally [18]. One of the findings of these early studies is that tumor growth is not exponential provided it is observed on a long enough time frame (100 to 1000 folds of increase) [13]. The specific growth rate slows down and this deceleration can be particularly well captured by the Gompertz model

[21, 13, 15]. The analytical expression of this model writes (where V_0 is the initial tumor size at $t = 0$ and α and β are two parameters):

$$V(t) = V_0 e^{\frac{\alpha}{\beta}\left(1 - e^{-\beta t}\right)} \tag{1}$$

While the etiology of the Gompertz model has been long debated [10], several independent researchers have reported a strong correlation between the parameters α and β estimated on distinct subjects within the same species [13, 16, 6]. While some suggested this would imply a constant maximal tumor size (given by $V_0 e^{\frac{\alpha}{\beta}}$ in (1)) across tumor types within a given species [6], others argued that because of the presence of the exponential, this could vary over several orders of magnitude [19]. To date, the generalizability, implications and understanding of this observation remain open questions in quantitative tumor growth.

Mathematical models for tumor growth have been previously studied at the level of individual kinetics and for prediction of future tumor growth [2]. However, up to our knowledge, a detailed study of statistical properties of classical growth models at the level of the population (*i.e.* integrating structural dynamics with inter-animal variability) remains yet to be reported. Longitudinal data analysis with nonlinear mixed-effect modelling provides an ideal tool for such a task [14]. In addition, the reduced number of parameters (from $p \times N$ to $p + \frac{p(p+1)}{2}$ where N is the number of animals and p the number of parameters of the model) ensures a higher robustness of the estimates, in the sense of smaller standard errors. Therefore, this framework is particularly adapted to study the above-mentioned correlation between the two Gompertz parameters.

Moreover, using the population distribution as prior allows to make predictions on new subjects by means of Bayesian algorithm such as the Hamiltonian Monte Carlo algorithm [12, 11], implemented in `Stan` [7]. The advantage of this method is that only few measurements of the new individual are necessary to have reliable prognosis.

2   Material and methods

Mice experiments. The experimental data consisted in murine Lewis lung carcinoma cells originally derived from a spontaneous tumor in a C57BL/6 mouse [4]. They were implanted subcutaneously (10^6 cells at injection) on the caudal half of the back in anesthetized 6- to 8-week-old C57BL/6 mice. Tumor size was measured as described for the breast data. The data was pooled from two experiments with a total of 188 observations. A precise description of the experimental protocol is reported elsewhere (see [2]).

Tumor growth models. At the time of injection ($t_0 = 0$), we assumed that all the animal tumor volumes within a group have the same volume V_0 (taken to be equal to the number of injected cells converted in mm^3) and denote by α the specific growth rate ($\alpha = \frac{1}{V}\frac{dV}{dt}$) at this time and volume.

We considered the exponential, logistic and Gompertz models [2]. The first two are respectively defined by:

$$V_E(t; \alpha) = V_0 \exp(\alpha t) \qquad \text{and} \qquad V_L(t; \alpha, K) \frac{V_0 K}{(V_0 + (K - V_0)e^{-\alpha t})}. \qquad (2)$$

In the logistic equation, K is a carrying capacity parameter.

The Gompertz model $V_G(t; \alpha, \beta)$ is characterized by an exponential decrease of the specific growth rate with rate β. The differential form thus reads:

$$\begin{cases} \dfrac{dV_G}{dt} = \left(\alpha - \beta \log \left(\dfrac{V_G}{V_0} \right) \right) V_G, \\ V_G(t = 0) = V_0. \end{cases} \qquad (3)$$

Note here that the initial condition also appears in the differential equation defining V_G. This is natural from our assumption that α is the specific growth rate at the injected volume V_0.

Population approach. Let N be the total number of subjects within the population and $\boldsymbol{Y}^i = \{y_1^i, ..., y_{n^i}^i\}$ the vector of longitudinal measurements of the animal i, where y_j^i is the observation of subject i at time t_j^i for $i = 1, ..., N$ and $j = 1, ..., n^i$ (n^i is the total number of measurements of individual i). We assumed the following statistical model

$$y_j^i = V(t_j^i; \boldsymbol{\theta}^i) + e_j^i, \qquad j = 1, ..., n^i, \quad i = 1, ..., N, \qquad (4)$$

where $V(t_j^i; \boldsymbol{\theta}^i)$ is the evaluation of one of the tumor growth models at time t_j^i, $\boldsymbol{\theta}^i \in \mathbb{R}^p$ is the vector of the parameters relative to the individual i and e_j^i the residual error model, to be defined later. We assumed that the individual parameters $\boldsymbol{\theta}^i$ follow a lognormal distribution that are therefore identified by

$$\log(\boldsymbol{\theta}^i) = \log(\boldsymbol{\mu}) + \boldsymbol{\eta}^i,$$

where $\boldsymbol{\mu}$ denotes the fixed effects and $\boldsymbol{\eta}^i$ denotes the random effects. The former are identical within the population while the latter are specific for each animal and follow a normal distribution $\boldsymbol{\eta}^i \sim \mathcal{N}(0, \boldsymbol{\omega})$ with mean zero and variance matrix $\boldsymbol{\omega}$.

We considered a combined residual error model e_j^i, defined as

$$e_j^i = \left(\sigma_1 + \sigma_2 f(t_j^i; \boldsymbol{\theta}^i) \right) \varepsilon_j^i,$$

where $\varepsilon_j^i \sim \mathcal{N}(0, 1)$ are the residual errors and (σ_1, σ_2) are the residual error model parameter.

In order to compute the population parameters, we maximized a population likelihood, obtained by pooling together all the data. Usually, this likelihood cannot be computed explicitly for nonlinear mixed-effect models. The optimization

procedure can be implemented using the stochastic approximation expectation minimization algorithm (SAEM) [14], implemented in `Monolix` [1].

From now on we denote by $\phi = \{\boldsymbol{\mu}, \boldsymbol{\omega}, \boldsymbol{\sigma}\}$ the set of the population parameters containing the fixed effects $\boldsymbol{\mu}$ and the random effects $\boldsymbol{\omega}$ of the parameters and the vector of error model parameters $\boldsymbol{\sigma} = [\sigma_1, \sigma_2]$.

Individual predictions: Bayesian inference. We considered the problem of predicting the age of the tumor of an animal based on three late measurements. We splitted the data set into two subgroups: a *training set*, used to learn the population parameters distribution, and a *test set*, to assess the performance of the prediction.

Let us assume that the set of the population parameters ϕ has been identified on a *training set* using the population approach. We used this information to make predictions for a new animal j in the *test set* considering only its last three measurements $\boldsymbol{y}^j = \{y_{n^j-2}^j, y_{n^j-1}^j, y_{n^j}^j\}$. The posterior distribution $\mathbb{P}(\boldsymbol{\theta}^j | \boldsymbol{y}^j, \phi)$ of the parameters $\boldsymbol{\theta}^j$ was then given thanks to the Bayesian approach [11]:

$$\mathbb{P}(\boldsymbol{\theta}^j | \boldsymbol{y}^j; \phi) = \mathbb{P}(\boldsymbol{\theta}^j; \phi)\mathbb{P}(\boldsymbol{y}^j | \boldsymbol{\theta}^j; \phi), \tag{5}$$

where $\mathbb{P}(\boldsymbol{\theta}^j; \phi)$ is the prior distribution of the parameters found with the nonlinear mixed effects modeling and $\mathbb{P}(\boldsymbol{y}^j | \boldsymbol{\theta}^j; \phi)$ is the likelihood. Then we computed the posterior predictive distribution of $\tilde{y}^j(u)$, with $u < t_{n^j-2}$ defined as

$$\mathbb{P}(\tilde{y}^j(u) | \boldsymbol{y}^j) = \int_{\boldsymbol{\theta}^j} \mathbb{P}(\tilde{y}^j(u) | \boldsymbol{\theta}^j; \phi)\mathbb{P}(\boldsymbol{\theta}^j | \boldsymbol{y}^j; \phi)d\boldsymbol{\theta}^j. \tag{6}$$

We draw realizations for (5) and for (6) using `Pystan`, a Python interface to the software `Stan` [7] for Bayesian inference based on the No-U-Turn sampler, a variant of Hamiltonian Monte Carlo [12]. These realizations were then used to estimate tumor growth kinetic as the median value of the sample.

3   Results

In [20] other two data sets (two animal models of breast cancer, measured by volume and fluorescence) are considered for the analysis with equivalent results.

3.1   Population analysis of tumor growth curves

We applied the population approach to test the descriptive power of the exponential, logistic and Gompertz models for tumor growth kinetics. The number of injected cells at time $t_0 = 0$ was 10^6, therefore we fixed the initial volume $V_0 = 1$ mm^3 in the whole dataset [2].

We ran the SAEM algorithm with the `Monolix` software to estimate the fixed and random effects. Moreover, different statistical indices were evaluated in order to compare the different tumor growth models. We report them in

Table 1, where the models are ranked according to their AIC (Akaike information criterion). As shown below, the Gompertz model provided the lowest AIC values. Different types of model diagnostic plots are reported in Figure 1. The visual predictive checks (VPCs) in Figure 1A compare the empirical percentiles with the theoretical percentiles, *i.e.* those obtained from simulations of the calibrated models. Only in the case of the Gompertz model the observed percentiles were close to the predicted ones and remained within the corresponding prediction interval. The VPCs of the exponential and the logistic models exhibited model misspecification. The observations vs individual predictions of the Gompertz model in Figure 1B show a low percentage of outliers, *i.e.* the predictions outside of the 90% prediction interval. Moreover, the distribution of the observations were symmetrical around the predicted values with the Gompertz model (Figure 1C) while the exponential and the logistic models provided skewed distributions. Figure 1D shows an example of individual fit with the three different models. This confirms that the Gompertz model describes better the dynamic of tumor growth.

Table 2 provides the values of the population parameters. The relative standard errors associated to population parameters were all low ($< 10.7\%$), indicating good practical identifiability of the model parameters. Relative standard errors of the standard deviations of the random effects $\boldsymbol{\omega}$ were all smaller than 34.1%.

Model	-2LL	AIC	BIC
Gompertz	2232	2246	2253
Reduced Gompertz*	+24	+20	+18
Logistic	+83	+81	+80
Exponential	+412	+406	+403

Table 1. Models ranked in ascending order of AIC (Akaike information criterion). Other statistical indices are the log-likelihood estimate (-2LL) and the Bayesian information criterion (BIC). The reported values in the first row are the values of the indices of the best model (the Gompertz model). The other rows provide the difference of each statistical index between the model in the row and the Gompertz model. *The reduced Gompertz model is introduced in Section 3.2.

3.2   The reduced Gompertz model

Correlation between the Gompertz parameters. Although the Gompertz parameters α and β were assumed to be independent, a high correlation within the population has been observed. Indeed, the SAEM algorithm estimated a correlation of the random effects equal to 0.957. Moreover, Figure 2A shows the relation between the individual parameters, where we found $R^2 = 0.929$. This correlation was observed in other experimental system: in a rat mammary

Model	Parameter	Unit	Fixed effects	CV (%)	R.S.E. (%)
Gompertz	α	day^{-1}	0.713	22.57	3.79
	β	day^{-1}	0.0731	318	5.77
	σ	-	[28.2,0.081]	-	[13.8,14.3]
Reduced Gompertz*	β	day^{-1}	0.0757	158.37	10.7
	k	-	9.51	-	5.26
	σ	-	[27.6,0.106]	-	[14.03,11.7]
Logistic	α	day^{-1}	0.477	25.48	2.84
	K	mm^3	1.65e+03	0.006	4.67
	σ	-	[38.5,0.11]	-	[13.2,14.01]
Exponential	α	day^{-1}	0.403	28.01	2.75
	σ	-	[87.8,0.37]	-	[19.1,14.8]

Table 2. Fixed effects (typical values) of the parameters of the different models. CV = Coefficient of Variation, expressed in percentage and estimated as the standard deviation of the parameter divided by the fixed effect and multiplied by 100. σ is vector of the residual error model parameters. Last column shows the relative standard errors (R.S.E.) of the estimates. *The reduced Gompertz model is introduced in Section 3.2.

carcinoma system [16], human lung metastases from testicular tumors [9] and human benign tumors [17]. We write the relationship between the two parameters as:

$$\alpha^i = k\beta^i + c \qquad (7)$$

where c is the intercept of the regression line, which is found close to zero. The slope of the regression line could be a characteristic constant of tumor growth within a certain species. From a biological point of view, this characteristic constant could be associated to the carrying capacity K, following the relation $K = V_0 \exp(k)$, where V_0 is the initial volume of the tumor. As previously remarked by [6], this result might be supported by the fact that a particular species is able to support a tumor of a certain maximum size.

Biological interpretation in terms of the proliferation rate. By definition, the parameter α is equal to the specific growth rate at the time of injection. Assuming that the cells do not change their proliferation kinetics when implanted, this value should thus be equal to the *in vitro* proliferation rate (supposed to be the same for all the cells of the same cell line), denoted here by λ. The value of this biological parameter was assessed *in vitro* and found equal to 0.929 [3]. Confirming our theory, we indeed found estimated values of α close to λ (fixed effects of 0.713), although strictly smaller in the majority of the cases (Figure2A). This difference could be explained by the fact that not all the cells "take" when grafted in an animal. Denoting by $\hat{V}_0^i < V_0$ the volume of these

cells, our assumption would rather be expressed as:

$$\lambda = \alpha^i - \beta^i \log\left(\frac{\hat{V}_0^i}{V_0}\right) > \alpha^i,$$

which was confirmed in our observations.

Population analysis of the reduced Gompertz model. The high correlation among the Gompertz parameters, combined to the biological rationale explained above, suggested that a reduction of the degrees of freedom could improve identifiability of the parameters and yield a simpler model. Considering the relation in (7), and assuming c negligible, we thus propose the following reduced Gompertz model $V_R(t; \beta, k)$:

$$\frac{dV_R}{dt} = \beta^i k - \beta^i \log\left(\frac{V_R}{V_0}\right), \qquad i = 1, ..., N \tag{8}$$

where β has mixed effects, while k has only fixed effects, *i.e.* k is constant within the population.

Figure 2 shows the results relative to the population analysis performed with `Monolix`. We noticed a good description of the population (Figure 2B) and of the individual trends (Figure 2C) even if only one parameter has mixed effects. Moreover, the residuals are symmetrically distributed around zero (Figure 2D). Table 1 shows the statistical indices of the 1-d Gompertz model. Comparing these values with the other equations we noticed that the reduced model performes well compared to the other growth curves. Moreover, we obtained an excellent identifiability of the parameters (Table 2).

3.3   Prediction of the time since tumor initiation

We then studied the relative performances of the reduced Gompertz and the Gompertz models for the problem of predicting the initiation time from the three last measurements using Bayesian inference. For a given animal i, we consider as first observation $y_{n^i-2}^i$ and tried to predict $t_{n^i-2}^i$. Initial conditions were not considered equal to the number of injected cells anymore but rather to $y_{n^i-2}^i$. The value t_{pred}^i was defined as the time when the median value of the posterior predictive distribution of $\tilde{y}(u)$ reached V_0.

Different data sets were used for learning the priors (*training sets*) and for making predictions (*test sets*) by means of k-fold cross validation, with k equal to the total number of animals of the dataset ($k = N$). At each iteration we computed the parameters distribution of the population composed by $N - 1$ individuals and used this as prior to predict the initiation time of the excluded subject i. The `Stan` software was used to draw 2000 realizations from the posterior predictive distribution of the animal i. We eventually estimated the model accuracy (i.e. relative error of the prediction, defined by $\text{err}^i = t_{\text{pred}}^i / t_{n^i-2}^i$) and the uncertainty of the prediction (i.e. precision, measured by the width of the 90% prediction interval (PI)).

Figure 3 shows some examples of prediction of three individuals and Figure 4 shows the distribution of the relative error. The reduced Gompertz model was found to have better accuracy in predicting the initiation time (mean error = 9.4%) and to have the smallest uncertainty (mean precision = 7.34 days), while the Gompertz model had worse performances (mean error = 19.6% and precision = 18.2 days). Indeed the reduced Gompertz had only one parameter to estimate and the prior distribution allowed to have a reliable prediction.

4   Conclusions

We have performed a quantitative analysis of tumor growth kinetics using nonlinear mixed-effects modeling. This allowed us to propose a novel, "reduced" Gompertz model with one parameter less. We further developed a method for prediction of individual tumor age given few measurements. The approach is based on: (i) the application of the population approach in order to learn the parameter distribution of the models, (ii) the reduced Gompertz model with only one individual parameter and (iii) Bayesian inference to determine the posterior predictive distribution used to compute the time since initiation.

Our results warrant against the use of the exponential or logistic models for description of tumor growth, that were therefore excluded in the prediction of the age of a tumor. On the other hand, combining the population approach with a reduced version of the Gompertz model comprising one parameter only allows to reach a level of accuracy which offers promising clinical perspectives.

The method that we propose here remains to be extended to clinical data, although it will not be possible to have a firm confirmation since the entire natural history of neoplasms cannot be observed. Nevertheless, the encouraging results obtained here could allow to give approximate estimates. Such predictions could be informative in clinical practice to determine the extent of invisible metastatics at the time of diagnosis, by refining published methods [5].

References

1. Monolix version 2018R2. Lixoft SAS (2018)
2. Benzekry, S., Lamont, C., Beheshti, A., Tracz, A., Ebos, J.M.L., Hlatky, L., Hahnfeldt, P.: Classical Mathematical Models for Description and Prediction of Experimental Tumor Growth. PLoS Comput Biol **10**(8), e1003800 (Aug 2014). https://doi.org/10.1371/journal.pcbi.1003800
3. Benzekry, S., Tracz, A., Mastri, M., Corbelli, R., Barbolosi, D., Ebos, J.M.L.: Modeling Spontaneous Metastasis following Surgery: An In Vivo-In Silico Approach. Cancer Res **76**(3), 535–547 (Feb 2016)
4. Bertram, J.S., Janik, P.: Establishment of a cloned line of Lewis Lung Carcinoma cells adapted to cell culture. Cancer Lett. **11**(1), 63–73 (Nov 1980)
5. Bilous, M., Serdjebi, C., Boyer, A., Tomasini, P., Pouypoudat, C., Barbolosi, D., Barlesi, F., Chomy, F., Benzekry, S.: Quantitative mathematical modeling of clinical brain metastasis dynamics in non-small cell lung cancer. Sci Rep **9**(1) (Dec 2019). https://doi.org/10.1038/s41598-019-49407-3

6. Brunton, G.F., Wheldon, T.E.: Characteristic species dependent growth patterns of mammalian neoplasms. Cell Tissue Kinet **11**(2), 161–175 (Mar 1978)
7. Carpenter, B., Gelman, A., Hoffman, M.D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P., Riddell, A.: Stan : A Probabilistic Programming Language. J Stat Softw **76**(1) (2017). https://doi.org/10.18637/jss.v076.i01
8. Collins, V.P., Loeffler, R.K., Tivey, H.: Observations on growth rates of human tumors. Am J Roentgenol Radium Ther Nucl Med **76**(5) (1956)
9. Demicheli, R.: Growth of testicular neoplasm lung metastases: Tumor-specific relation between two Gompertzian parameters. Eur J Cancer **16**(12), 1603–1608 (Dec 1980). https://doi.org/10.1016/0014-2964(80)90034-1
10. Frenzen, C.L., Murray, J.D.: A Cell Kinetics Justification for Gompertz' Equation. SIAM J Appl Math **46**(4), 614–629 (Aug 1986)
11. Gelman, A.: Bayesian Data Analysis. Chapman & Hall/CRC Texts in Statistical Science, CRC Press, Boca Raton, third edition edn. (2014)
12. Kramer, A., Calderhead, B., Radde, N.: Hamiltonian Monte Carlo methods for efficient parameter estimation in steady state dynamical systems. BMC Bioinformatics **15**(1), 253 (2014). https://doi.org/10.1186/1471-2105-15-253
13. Laird, A.K.: Dynamics of tumor growth. Br J Cancer **13**, 490–502 (Sep 1964)
14. Lavielle, M.: Mixed Effects Models for the Population Approach: Models, Tasks, Methods and Tools. Chapman & Hall/CRC Biostatistics Series, Taylor & Francis, Boca Raton (2014)
15. Norton, L.: A Gompertzian model of human breast cancer growth. Cancer Res **48**(24), 7067–7071 (Dec 1988)
16. Norton, L., Simon, R., Brereton, H.D., Bogden, A.E.: Predicting the course of Gompertzian growth. Nature **264**(5586), 542–545 (Dec 1976). https://doi.org/10.1038/264542a0
17. Parfitt, A.M., Fyhrie, D.P.: Gompertzian growth curves in parathyroid tumours: Further evidence for the set-point hypothesis. Cell Prolif. **30**(8-9), 341–349 (1997 Aug-Sep)
18. Steel, G.G.: Growth kinetics of tumours: cell population kinetics in relation to the growth and treatment of cancer. Clarendon Press (1977)
19. Steel, G.G.: Species-dependent growth patterns for mammalian neoplasms. Cell Tissue Kinet **13**(4), 451–453 (Jul 1980)
20. Vaghi, C., Rodallec, A., Fanciullino, R., Ciccolini, J., Mochel, J., Mastri, M., Poignard, C., Ebos, J.M., Benzekry, S.: A reduced Gompertz model for predicting tumor age using a population approach. bioRxiv (Jun 2019). https://doi.org/10.1101/670869
21. Winsor, C.P.: The Gompertz curve as a growth curve. Proc Natl Acad Sci U S A **18**(1), 1–8 (1932)
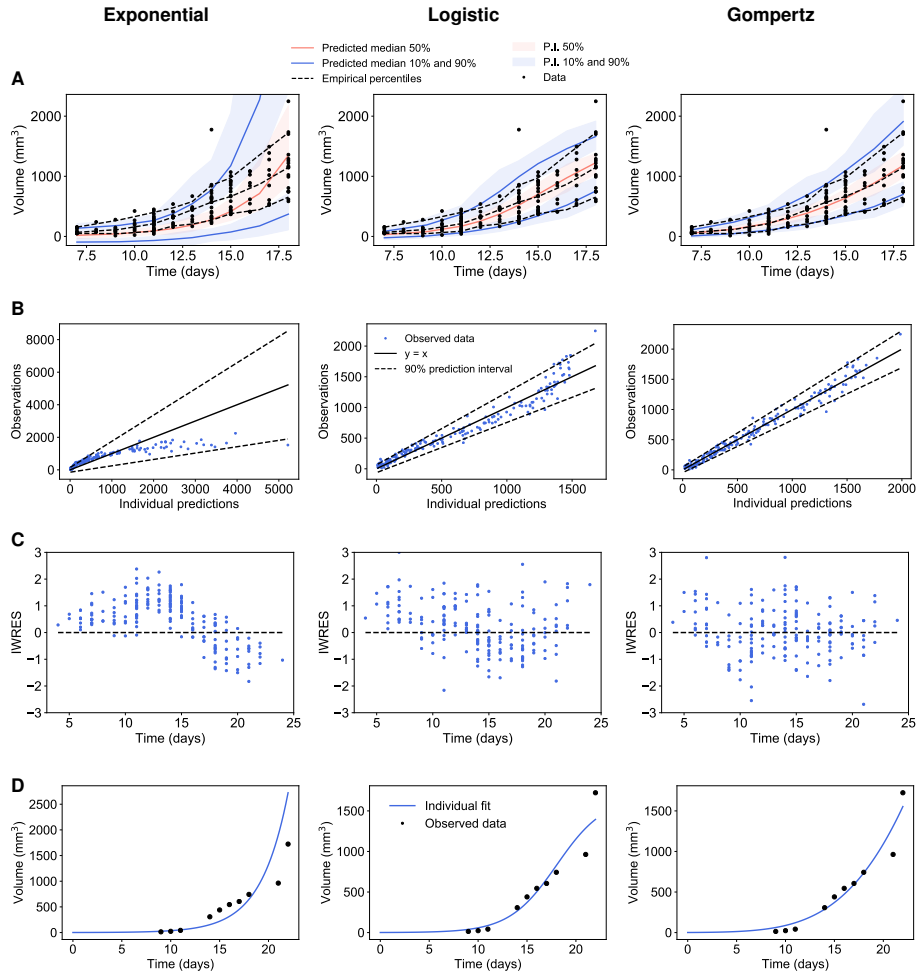
Fig. 1. Results relative to the population analysis: visual predictive check (A), observations vs predictions (B), distribution of the individual weighted residuals with respect to time (C) and example of an individual fit (D) of the exponential (left), the logistic (center) and the Gompertz (right) models.
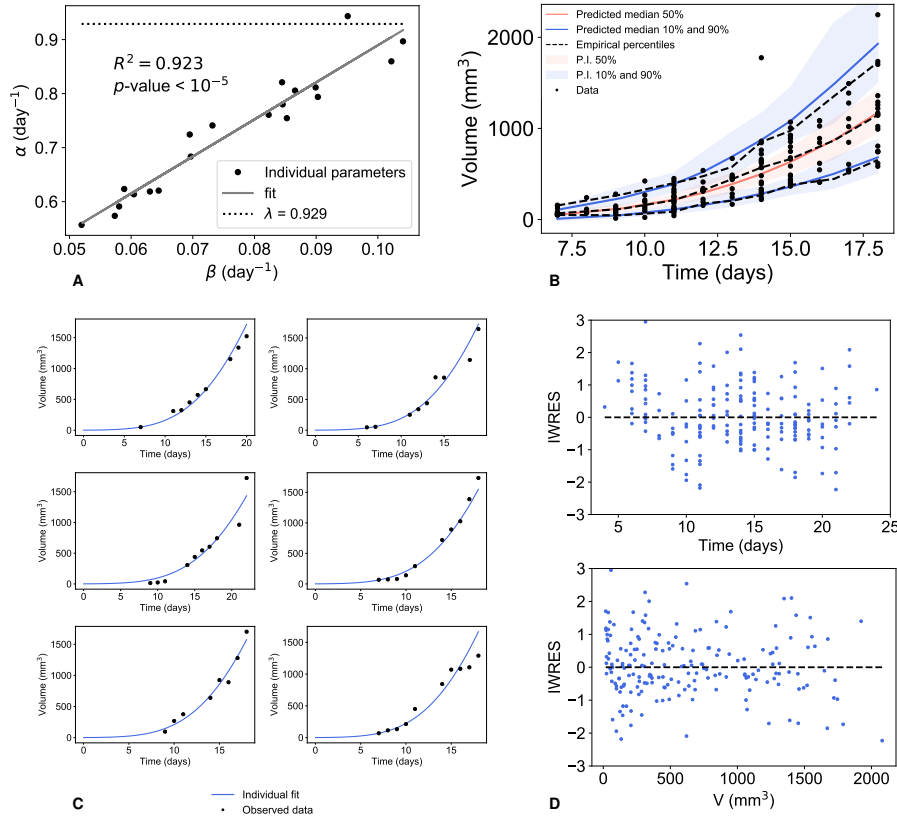
Fig. 2. Correlation between the parameters of the Gompertz model (A) and results of the population analysis of the reduced Gompertz model : visual predictive check (B), examples of individual fits (C) and scatter plots of the residuals (D).
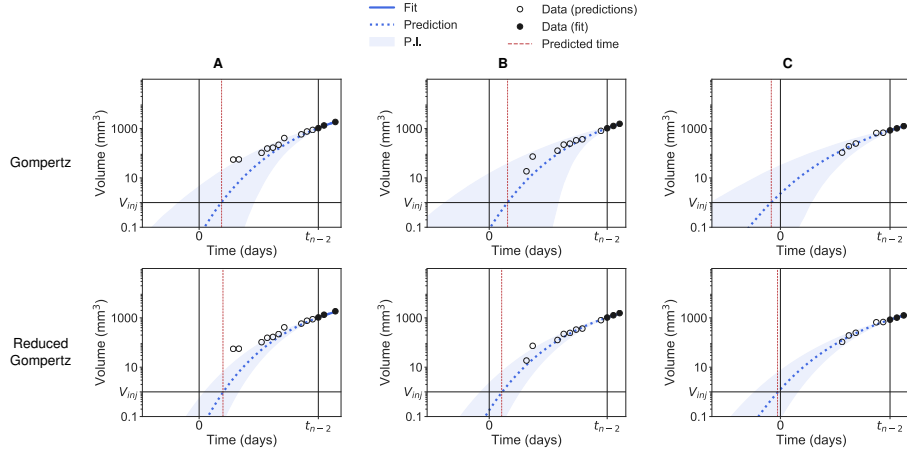
Fig. 3. Three examples of backward predictions of individuals A, B and C computed with Bayesian inference: Gompertz model (first row) and reduced Gompertz (second row). Only the last three points are considered to estimate the parameters. The grey area is the 90% prediction interval (P.I) and the dotted blue line is the median of the posterior predictive distribution. The red line is the predicted initiation time and the black vertical line the actual initiation time.
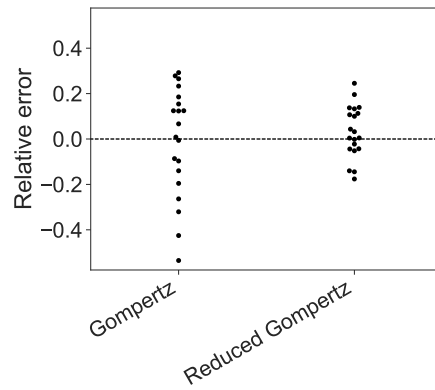

Fig. 4. Accuracy of the prediction models: swarmplots of relative errors of the Gompertz and the reduced Gompertz model.