

Letters to the Editor

Notes on reparameterization of bacterial growth curves

In representing microbial growth response, the logarithm of the number of viable cells per unit (logarithm of the cell concentration) is commonly plotted against the time elapsed. It is desirable to apply mathematical curve fitting procedures so that growth parameters can be determined objectively under the particular conditions of culture.

The basic step in the mathematical abstraction is the assumption that the main factors of the environment such as temperature, pH, water activity, etc. unambiguously determine the responses of particular bacteria (Frederickson et al. 1967). This is equivalent to the assumption that the parameters of a selected growth curve are defined by these environmental factors. Therefore it is crucial to find a sound basis for fitting growth curves to data obtained by a growing bacterial population.

In order to describe $\ln N_t$, the logarithm of the concentration of a growing bacterial culture, Gibson et al. (1988) selected the Gompertz curve with an additive parameter. Following the notation of Garthright (1991), this four parameter family of curves has the form

$$G_4(t) = D + C \exp \{-\exp [-B(t-M)]\}. \quad (1)$$

Zwietering et al. (1990) suggested the so-called 'modified' Gompertz equation by considering the three parameter curve.

$$g_3(t) = C \exp \{-\exp [-B(t-M)]\} \quad (2)$$

in the reparameterized form

$$g_3(t) = a \exp \{-\exp [e\mu\alpha^{-1}(L-t) + 1]\} \quad (3)$$

for the quantity $\ln (N_t/N_0)$. Garthright (1991) reports that Eqn (3) is a special case of Eqn (1) when the additive

parameter D is estimated by $\ln N_0$. Therefore it would be more precise to call Eqn (3) a 'restricted' rather than a 'modified' Gompertz equation.

Note that the term 'Gompertz model' would not be very appropriate because it is used in the case when the per capita rate of a growing population (the specific growth rate here) is monotone decreasing according to the Gompertz law (Holgate 1989). Equations (1) and (3), however, can be used as empirical relations for the logarithm of the cell concentration to fit data.

Ratkowsky (1983) used the terminology model functions to denote different representations obtained by reparameterization of the same model. It is important to bear in mind that reparameterization does not change the model itself, but it changes the estimating procedure, called estimator, of the model parameters. Consequently, the aim of a reparameterization is to obtain a statistically more reliable estimator for the parameters in question.

Garthright (1991) points out that if the additive term D is to be estimated *a priori* then the choice $D = \ln N_0$, as applied in Zwietering et al. (1990), is biased because this estimates the lower asymptote of the growth curve by a value which should be ideally (in an error-free case) over the asymptote. He proposes to return to the four parameter curve (1) to fit $\ln N_{t1}$ but with the parameterization of (3):

$$g_4(t) = D + a \exp \{-\exp [e\mu\alpha^{-1}(L-t) + 1]\}. \quad (4)$$

To support this reparameterized form of the Gompertz equation he argues that among others:

- (a) the parameters to be estimated are those which are the most important from biological point of view (this point is emphasized also by Zwietering et al. 1990);
- (b) starting from different inoculum levels, the parameters μ and L of equation (4) are more or less constant while this does not hold for the parameters of Eqn (1).

According to Ratkowsky (1983) the main reasons why we look for different model functions (different function-representations) of a model, are:

- (a) to decrease the bias of the estimator;
- (b) to decrease the so-called parameter-effect-non-linearity;
- (c) to obtain an estimator which results in a closer-to-normal distribution of the estimated values;

A well-known fourth reason is:

- (d) to decrease the numerical instability of the estimator (characterizable for example by the so-called condition number — see Stoer and Bulirsch 1981).

Ratkowsky (1983) also mentions that to obtain parameters with accepted biological or physical meanings is not an important objective of reparameterization; in fact 'meaningful' parameters are frequently transformed to get a form which has better statistical and numerical features. One set of the parameters can be transformed into the other by the formula of the reparameterization which implies that either in finding initial estimates for the parameters or interpreting the results of the estimator or calculating the respective correlations, the formula of reparameterization can make the two forms similar as far as their easy-to-use properties are concerned. The fact that a model function, say F_1 , contains 'meaningful' parameters, and another model function of the same model, say F_2 , does not, has no

significant role in the statistical, numerical comparison of F_1 and F_2 .

To study the properties of the different model-functions of the same model, Ratkowsky (1983) emphasizes the importance of simulation studies when the exact parameter values are known and the model-error (the fact that a model never describes nature exactly) has no influence on the results.

Garthright (1991) uses real data examples to prove that one parameterization of a model (one model function) is better than another. This is in contradiction with the principle that different model functions of a given model should be compared independently of the question whether the model itself fits or not.

For similar reasons, the fact that some parameters of Eqn (4) are more or less constant for different inocula (Garthright 1991) does not prove that its parameterization is better than that of the Eqn (1).

After simulation studies on a set of typical values of the parameters and on 1000 computer-generated data sets as suggested by Ratkowsky (1983), I have found that there is no significant difference between the two forms, (1) and (4), of the Gompertz equation. [In fact the form used by Gibson et al. (1988) showed slightly better statistical properties.]

It is important to emphasize again that the question 'which model is better?' should be examined with real data, but the investigation 'which parameterization of a given model is better?' should be independent of real data, otherwise misleading results can be obtained.

References

- Frederickson, A. G., Ramkrishna, D. and Tsuchiya, H. M. (1967) Statistics and dynamics of procaryotic cell populations. *Math. Biosci.* **1**, 327–374.
- Garthright, W. E. (1991) Refinements in the prediction of microbial growth curves. *Food Microbiol.* **8**, 239–248.
- Gibson, A. M., Bratchell, N. and Roberts, T. A. (1988) Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride and storage temperature. *Int. J. Food Microbiol.* **6**, 155–178.
- Holgate, P. (1989) Variates of a stochastic model: a comparative study of the Gompertz effect. *J. Theor. Biol.* **139**, 369–378.
- Ratkowsky, D. A. (1983) *Nonlinear regression modelling*. New York, Marcel Dekker.
- Stoer, J. and Bulirsch, R. (1981) *Introduction to numerical analysis*. Berlin, Springer – Verlag.
- Zwietering, M. H., Jongenburger, I., Rombouts, F. M. and van't Riet, K. (1990) Modelling of the bacterial growth curve. *Appl. Environ. Microbiol.* **56**, 1875–1881.

Notes on reparameterization of bacterial growth curves — A reply to J. Baranyi

J. Baranyi's concern for the best parameterization of the Gompertz formula for log growth of bacterial populations is very welcome. Although his simulation studies 'found no significant difference between' the (B, M) form and the (L, μ) form, I am interested to pursue even what he calls the 'slightly better statistical properties' of the (B, M) form and hope he will publish or otherwise communicate to me the specifics of his simulations. I did not, however, argue from biological meaningfulness nor use 'real data' to distinguish statistically the two parameterizations. Nor did I claim improved statistical properties such as parameter-induced non-linearity. These are straw men that Baranyi demolishes. My argument for using (L, μ) is found in my subsection 'Lag time and growth rate' on page 241 of the article. My recommendation of (L, μ) is based on considerations other than statistical, stemming from its insensitivity to initial inoculum levels.

I might have added, however, that some microbiologists conducting these studies were not (and will not be) regressing a separate D value from each specific environment but were fixing D at their estimated $\log(N_0)$. This tended to stabilize B and M estimates, but with some small loss of accuracy. The strong imposition of a fixed D would not be necessary with the (L, μ) form. On the other hand, those who use (B, M) and did (or will) regress D on the growth data obviously would not regress D on the environmental variables. Therefore, these D values will be lost for the purpose of applications software.

The applications software of Buchanan uses regression equations on environ-

mental parameters to predict B and M ; then it uses a single representative N_0 of the original experiments to calculate L and μ ; and finally it applies the (L, μ) form to the user's (different) N_0 . Some small distortion is inevitable here since only a single value is used to represent the many slightly different experimental inocula that influenced the B s and M s.

I omitted the complexities above from the article because I thought the non-statistical reason given in the article was sufficient justification: i.e. comparison (between investigators) of the formulas that predict the growth curve parameters from the environmental parameters will be easier if everyone uses parameters that are less sensitive to differing densities of the inocula. In addition, the complexities cited above show that application of the models will be more straightforward and will have one less source of possible distortion.

Notes on reparameterization of bacterial growth curves — A reply to J. Baranyi and W. E. Garthright

The Gompertz equation was first used for microbiological modelling by Gibson et al. (1988) in the form

$$\log(N_t) = \log(N_0) + C \exp\{-\exp[-B(t-M)]\}. \quad (1)$$

Where t is the time since inoculation; N_t is population density at time t and N_0 at time 0; C is the population growth from inoculation to the stationary phase; B a function of maximum growth rate and C ; and M is the time of the inflection point of the function. The maximum growth rate is BC/e and lag time is $(M - 1/B)$. Predictive models would require equations to calculate specific values for M , B and the maximum population [$\log(N_0) + C$].

Zweitering et al. (1990) reparameterized the equation and Garthright (1991) modified it to

$$\log(N_t) = D + A \exp\{-\exp[e\mu(L-t)/A + 1]\}. \quad (2)$$

L is the lag time; μ maximum growth rate; A is the log population growth from inoculation to the stationary phase; D is calculated from the stationary phase population (U) less growth (A); the inoculum is the value of this expression at time t_0 , not the value of D . Predictive models would require values for L , μ and U .

The advantage claimed by Garthright (1991) for Eqn (2) is that μ and L are independent of inoculum size, whereas M and B in Eqn (1) are not. This would simplify comparisons between growth curves having different initial populations. In addition, data fitted using the modified form [Eqn (2)] should estimate the inoculum size more accurately than Eqn (1).

Baranyi challenged Garthright's assertion that Eqn (2) is better than the original [Eqn (1)] stating that demonstrating superiority of one reparameterized model over another should be shown using a series of simulated data sets where individual data points are given randomized variation. Real data should be used to evaluate different models but not to compare two forms of the same model.

From the purely mathematical/statistical perspective Baranyi is correct. However, I view microbiological models as a predictive tool; all are simplifications of the 'real world' and all are approximations. An old cliché says all models are incorrect, but some are useful. It would be instructive to have Baranyi's comparison of the two forms and know what the 'slightly better statistical properties' of Eqn (1) actually were. Likewise, demonstration that the μ and L parameters are independent of inoculum size is needed before that aspect of Eqn (2) can be relied upon. Realistically, the inherent variation in the collection of microbial growth data places a practical limit on the refinement of any model. Also, a comparison of different models against a set of real growth data cannot conclusively demonstrate which model is superior, there may be untried conditions where one model would be preferred over the others. I do not believe there is a substantial difference between either form of the Gompertz equation and recommend that a modeller employ the form judged to be most appropriate for his particular study.

I think it is constructive to recognize the lineage of the Gompertz equation;

B. Gompertz (1825) used the function to express human mortality. It was adopted by microbiologists because it appeared to fit typical growth curves. Gibson et al. (1988) and Zwietering et al. (1991) demonstrated the Gompertz fit better than logistic or other sigmoidal functions. Garthright (1991) pointed out several intrinsic properties of the equation, in particular, that the time of the maximum growth rate (M value) is the time when the culture has grown 37% from the logarithm of the inoculation to stationary phase and the end of the lag phase is the time when growth reaches 6.6% of the logarithm of the total growth. In addition, the curvatures of the initial and final phases of exponential growth have a defined asymmetric relationship between each other. I do not believe microbiologists would be

comfortable having any of these properties of the Gompertz function construed as inherent characteristics of microbial growth.

A more profitable endeavor would be to create mechanistic- or kinetic-based models derived from observations and assumptions about microbial growth. Testing these models against microbial growth data would more likely lead to advances in microbial modelling than additional refining of an empirical model.

References

- Garthright, W. E. (1991) Refinements in the prediction of microbial growth curves. *Food Microbiol.* **8**, 239–248.
- Gibson, A. M., Bratchell, M. and Roberts, T. A. (1988) Predicting microbial growth: growth responses of salmonellae in a laboratory medium as affected by pH, sodium chloride and storage temperature. *Int. J. Food Microbiol.* **6**, 155–178.
- 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 contingences. *Philos. Trans. R. Soc. London* **115**, 513–585.
- Zwietering, M. H., Jongenburger, I., Rombouts, F. M. and van't Riet, K. (1990) Modelling of the bacterial growth curve. *Appl. Environ. Microbiol.* **56**, 1875–1881.