

Modelling and parameter estimation  
of bacterial growth with distributed  
lag time

PhD Thesis

by

József Baranyi

Doctoral School of Informatics  
University of Szeged, Hungary

2010

## Most frequently used variables and parameters

- $\alpha(t)$  A quantity characterising the physiological state of the bacterial culture at the time  $t$  ( $0 \leq \alpha(t) \leq 1$ ). In simplest form,  $\alpha(t) = \mu(t) / \mu_{max}$ .
- $\alpha_0$  Value of  $\alpha(t)$  (at the moment of inoculating a bacterial culture in a new environment).  $\alpha_0 = e^{-\mu \cdot \lambda}$
- $\alpha_N$  (Initial) Physiological state parameter for a population consisting of  $N$  cells at inoculation.
- $h_0 = \mu \cdot \lambda = -\ln \alpha_0$
- $\lambda$  Generic notation for the lag time (defined in various, empirical or mechanistic ways).
- $L, L_g$  In the exponential phase, the cell population grows, on the log scale, as the delayed linear function  $y(t) = \mu \cdot (t - L)$ . The  $_{-g}$  subscript refers to the fact that this is a rather “geometrical” (opposed to “physiological”) definition.
- $L^{(i)}$  The  $L$ -lag for the cell  $i$ . Though strictly speaking  $\tau_i$ , the physiological lag, is not the same as  $L^{(i)}$  we identify them unless their very difference is studied. Numerical simulation can prove that  $E(L^{(i)}) = E(\tau_i)$  and their variance is close to each other, provided the generation times in the exponential phase do not have too big variance.
- $L_N$  “Geometrical” lag as above, emphasizing that it is generated by  $N$  initial cells.
- $\mu(t)$  Instantaneous specific growth rate at the time  $t$ :  
 $\mu(t) = dx/dt / x(t) = d(\ln x(t)) / dt$
- $\mu_{max}, \mu$  (Maximum) specific growth rate

$m$	Curvature parameter for our sigmoid growth model, characterizing the transition from the exponential to the stationary phase.
$V_b$	Specific accumulation rate of a bottle-neck intracellular substance needed to overcome the lag time
$n_C$	Curvature parameter characterizing the transition from the lag to the exponential phase. In our model $n_C = V_b / \mu_{max}$
$N, N_0$	Inoculation level of a batch culture. $x_0$ is also used.
$\tau_i$	A random variable, the “physiological” lag time of the single cell $i$ . The time to the first division is the sum of $\tau_i$ and the first (random) generation time.
$t$	time
$T_{det}$	Detection time. $x(T_{det}) = X_{det}$
$x(t)$	Cell concentration at the time $t$
$x_0$	Initial cell concentration, $x(0)$
$x_{max}$	Maximum population density; the carrying capacity of the environment.
$X_{det}$	Detection level (constant): $x(T_{det}) = X_{det}$
$y(t)$	Natural logarithm of cell concentration at the time $t$ : $y(t) = \ln x(t)$
$y_0$	Natural logarithm of initial cell conc.: $y_0 = y(0)$
$y_{max}$	Natural logarithm of maximum population density: $y_{max} = \ln X_{max}$
$y_{det}$	Natural logarithm of the detection level: $y_{det} = \ln X_{det}$

## Introduction

In the 80's, with powerful desktop computing becoming everyday use, the name 'predictive microbiology' was coined for an area of food microbiology. Its closest relative can be found in biotechnology with its predictive mathematical models well elaborated in the 60's.

At the beginning of this thesis, we found major differences in requirements for mathematical models applied to biotechnology problems and those to be developed for food microbiology. New approaches were needed for modelling bacterial growth and survival in food.

### Deterministic modelling for bacterial population

Let  $x(t)$  describe the variation of the size of the bacterial population with the  $t$  time. In a constant environment, bacterial growth is usually described by a linear differential equation with a constant coefficient:

$$\frac{dx}{dt} = \mu x$$

Here  $\mu$  is called the specific growth rate of the population. Its value is characteristic to the species and the environment. It is close to the average number of divisions of a cell per unit time. It is a function of environmental factors such as temperature, pH etc. It can be estimated by fitting a linear function to the exponential phase of the "ln  $x(t)$  v. time" curve.

In food microbiology laboratories, typically batch cultures of bacteria are generated, that form a sigmoid curve on the log scale. The phase before the exponential (on log scale: linear) phase is called the adaptation/adjustment period, or lag phase. While this phase is ignored in biotechnology research, it is of

paramount importance in food microbiology and Microbial Risk Assessment.

Sigmoid curves can be obtained, on the arithmetical scale by means of the Initial Value Problem

$$\frac{dx}{dt} = \mu(x) x$$
$$x(0) = x_0 \quad (0 < x_0 < x_{max} ; \quad x_{max} \text{ is fixed})$$

where

- $\mu: (0, x_{max}] \rightarrow \mathbb{R}$
- $\mu(x)$  is continuously differentiable on  $(0, x_{max})$
- $d\mu/dx$  is strictly negative on  $(0, x_{max})$
- $\mu(x_0) > 0$  and  $\mu(x_{max}) = 0$

Certain sigmoid curves, special cases of the above generic model, were in common use, until the 90-s, to describe population kinetics on the *log scale*. This was a purely empirical approach and, from mechanistic point of view, became untenable as the demand for sophistication in mathematical modelling increased also in food microbiology.

### Stochastic modelling for single cell

Let the initial number of a growing cell population be  $N_0=N$ . Omit the stationary phase. The end of the lag phase ( $L_N$ ) will be defined as the breakpoint of the bi-phasic linear function (where there is no growth in the first phase) fitted to the  $y(t)=\ln x(t)$  growth curve. This *population lag* is generated by  $N$  initial cells. Its mathematical relationship with the distribution of the lag time of individual cells is far from straightforward. In addition, it is hard to acquire information on the variability of single cells during the lag time because, in that phase, the cell concentration is small and it is complicated and expensive to introduce an automated system that would detect cells at all.

## Objectives

We set out to create new models for bacterial kinetics in food, from deterministic models for population growth, with their numerical aspects and implications, to stochastic models for the kinetics of individual cells, including their numerical algorithms for parameter estimations. Our aim was to help better understanding and predicting of bacterial responses to food environments and to improve Quantitative Microbial Risk Assessment in Food.

For the deterministic (population level) problems, we wanted to show where it was most important to replace empirical approaches with mechanistic approaches. We aimed at developing easy-to-implement models and analyse their numerical aspects.

With the stochastic (single cell level) problems, our objective was to develop formulae linking the distribution of single cell lag times and that of the population. The theoretical work would imply numerical algorithms to reconstruct the distribution of single cell's kinetic parameters from data at population level acquired by automated measurements

## Results

### Deterministic, non-autonomous models for population growth and their implementation

1. We applied a so-called adjustment function, to the  $\mu(x)$  function of the deterministic model described in the Introduction.

$$\frac{d}{dt} x = \alpha(t)\mu(x) \cdot x \quad (0 \leq t < \infty; \quad 0 < x)$$
$$x(0) = x_0 \quad (0 < x_0 < x_{\max})$$

where  $\alpha(t)$  is sufficiently smooth and may depend on the pre-inoculation environment; besides

$$0 \leq \alpha(t) \leq 1 \quad (0 \leq t < \infty)$$

$\alpha(t) \rightarrow 1$  monotone increasingly as  $t \rightarrow \infty$

The solution  $g_\alpha(t)$  for the obtained non-autonomous model can be expressed by means of the solution of the autonomous model,  $f$ , and the adjustment function,  $\alpha(t)$ . Its form is:

$$g_\alpha(t) = f(A(t)) \quad \text{where} \quad A(t) = \int_0^t \alpha(s) ds$$

This is a result of Baranyi et al (1993b) and this is the basis of all the specific models developed for user-friendly applications.

**2.** The use of the Hill-function proved to be one possible option for  $\alpha(t)$  – see Baranyi et al (1993a). The derived formulae were algebraic, but too complex for example to build them in an ordinary spreadsheet. Besides,  $\alpha(t)$  was not given in a mechanistic way, and it was dubious how to interpret it if the environment changed dynamically during lag.

**3.** Baranyi and Roberts (1994) introduced an adjustment function based on the Michaelis-Menten kinetics, which is until today the most successful for practical modelling. The result is:

$$\frac{d}{dt} x = \frac{q(t)}{1 + q(t)} \mu(x) \cdot x$$

$$\frac{d}{dt} q = v_b \cdot q$$

$$x(0) = x_0 \quad ; \quad q(0) = q_0$$

where  $v_b$  describes the speed of adaptation. As it turns out

$$\alpha(t) = \frac{q_0}{q_0 + e^{-v_b \cdot t}}$$

and

$$A(t) = t + \frac{1}{\nu} \ln \left( \frac{e^{-\nu_b t} + q_0}{I + q_0} \right)$$

Besides

$$\lambda = \frac{\ln(1+1/q_0)}{\nu_b}$$

is a good definition for the lag time:  $f(A(t))$  will be a sigmoid curve on the log scale and converges to  $f(t-\lambda)$ . The  $0 \leq \alpha_0 \leq 1$  parameter was suggested to quantify the initial physiological state of the population.

**4.** If the Richards-model is taken for the autonomous ODE

$$\frac{dx(t)}{dt} = \mu_{\max} x(t) \left( 1 - \left( \frac{x(t)}{x_{\max}} \right)^m \right)$$

then we obtain an algebraic solution and the natural logarithm of the cell concentration will vary with time as

$$y(t) = y_0 + \mu_{\max} t - \frac{1}{m} \ln \left( I + \frac{e^{m\mu_{\max} t} - 1}{e^{m(y_{\max} \cdot y_0)}} \right)$$

Using the above  $A(t)$  function instead of  $t$ , as suggested in points 1 and 3, we get a versatile sigmoid function, with six parameters, with two curvature parameters,  $m$  and  $n_C = \nu_b / \mu_{\max}$ . We introduced several reparameterization for the model for numerical stability; see Baranyi et al (1994, 1995a).

**5.** Assuming that the rate of adjustment,  $\nu_b$  and the potential maximum specific growth rate instantaneously takes up the value that is characteristic of the **E** actual environment:



$$\frac{dx(t)}{dt} = \frac{q(t)}{1+q(t)} \mu(\mathbf{E}(t))x \cdot u(x)$$

$$\frac{dq(t)}{dt} = v_b(\mathbf{E}(t))$$

This dynamic model can be used to predict bacterial growth even if the environment changes during the lag. Important simplifications were that  $v_b = \mu_{max}$  and  $q_0 = \text{const}$  were assumed for the different environments. These simplifications are based on regressions carried out using large datasets. The assumptions are certainly justified, if the dynamic environment is identified with the variation of the temperature, inside a strongly growth-supporting region, with not too abrupt changes (Baranyi et al, 1995).

*Stochastic model to describe the lag time of individual cells*

**6.** Here we consider only on the lag and growth phases.

Measure the lag ( $L_N$ ) by means of the fitted bi-phasic linear function generated by  $N$  initial cells:

$$y(t) = y_0 + \max(\mu(t-L_N), 0)$$

$$y_0 = \ln N$$

We call the obtained  $L$ -value as the “geometrical” lag, due to its definition. If  $N=1$ , we arrive at the single cell lag time,  $L_1$ . It is not the same as the physiological lag time, but for practical purposes the  $L$ -definition is more useful. For historical reasons we denote the  $L_i$ -value for the cell  $i$  also by  $\tau_i$ . Its expected value is  $\tau$ .

Assume that  $\tau_i$  ( $i = 1 \dots N$ ) are identically distributed random variables, with  $\tau$  as their expected value. Introduce the  $\alpha_i = e^{-\mu \tau_i}$  variables. Their arithmetical mean is closer and closer to the respective parameter of the whole population, as  $N$  is increasing:

$$\alpha(N) = e^{-\mu L_N} = \frac{\sum_{i=1}^N e^{-\mu \tau_i}}{N}$$

The physiological state parameter of the population turned out to be the arithmetical average of the individual cells' physiological state parameters. This implies:

$$\alpha(N) \xrightarrow{N \rightarrow \infty} \alpha \quad \text{and} \quad L_N \xrightarrow{N \rightarrow \infty} \frac{-\ln(\alpha)}{\mu} = \lambda$$

Furthermore, the rate of convergence can be estimated by:

$$\text{Var}[\alpha(N)] = \frac{\text{Var}(\alpha_i)}{N}$$

These results, published in Baranyi (1998), solved our basic problem and gave bases for further developments.

**7.** We compared a deterministic compartment-model with our stochastic model and proved that the results are the same if the bacterial division follows Poissonian birth process. For example,

$$L_N \xrightarrow{N \rightarrow \infty} \frac{\ln(1 + \mu \cdot \tau)}{\mu} = \lambda$$

can be obtained from compartmental models, too (Baranyi, 1998).

**8.** We drew parallels between the survival curves with “shoulders” and growth curves with lag, both at single cell and population levels. We analysed three different distributions for the individual survival/lag times and derived formulae to express some of their analytical properties.

The gamma distribution proved to have the best properties. For the growth situation, a generalisation of the formulae of the previous point can be obtained:

$$\alpha = E[\alpha_i] = \left(\frac{r}{r+1}\right)^p ; \quad \text{Var}[\alpha_i] = \left(\frac{r}{r+2}\right)^p - \left(\frac{r}{r+1}\right)^{2p}$$

where  $r = \kappa/\mu$  and the parameter  $\kappa$  is the scale-,  $p$  is the shape- parameter of the distribution. This result from Baranyi and Pin (2001) is the basis for the numerical methods described in the last chapter to measure the distributions of individual lag times.

We proved the following generalisation between the limit population lag,  $\lambda$ , and the mean individual lag time:

$$\lambda = \frac{\ln\left(1 + \frac{\mu}{p} \tau\right)}{\mu / p}$$

where  $\tau = p/\kappa$  is the mean of the gamma distribution. This is a generalisation of the result of point 7 and shows how the population lag depends on the shape of the distribution of the individual lag times.

**9.** Using the above formulae, we developed methods to estimate the expected value and variance of individual lag times. We assumed that experiments started from different initial levels where the ratio between the inocula was known (for example the lower was obtained from the higher by dilution), and that a culture starting from the level  $x_0$ , reaches the detection level in its exponential phase. Then we provided an ANOVA protocol:

Let  $T_{det}$  be the detection time when  $x(T_{det}) = X_{det}$ . Use notations as follows: we study  $m$  groups of cultures characterized by  $m$  different inoculum levels. In the group  $k$  ( $k=1 \dots m$ ) we have  $j=1 \dots n_k$  cultures. For a fixed  $j$ , the  $X_{det}$  detection level is reached at the time  $T^{(j)}$ , starting from the inoculum level  $x_0^{(j)}$ . Denote

the physiological state parameter of this culture by  $\alpha^{(j)} = \alpha(x_0^{(j)})$ .  
Then

$$\alpha^{(j)} = e^{-\mu T^{(j)}} / r^{(j)}$$

where  $r^{(j)} = x_0^{(j)} / X_{\text{det}}$ . Besides

$$\text{Var}(\alpha^{(j)}) = v / X_{\text{det}} r^{(j)}$$

where  $v$  is the common variance of the  $e^{-\mu T_i}$  individual  $\alpha_i$  physiological states. The grand mean of all  $\alpha_i$  is

$$\bar{\alpha} = \frac{\sum_{j=1}^n r^{(j)} \alpha^{(j)}}{\sum_{j=1}^n r^{(j)}} = \frac{\sum_{j=1}^n e^{-\mu T^{(j)}}}{r_{\text{sum}}} \quad \text{where} \quad r_{\text{sum}} = \sum_{j=1}^n r^{(j)}$$

The group means are :

$$\bar{\alpha}^{(k)} = \frac{\sum_{j=1}^{n_k} \alpha_{k,j}}{n_k} = \frac{\sum_{j=1}^{n_k} e^{-\mu T_{k,j}}}{n_k r^{(k)}}$$

Define the variance ratio

$$V = \frac{\sum_{k=1}^m \frac{r^{(k)} \left( \alpha^{(k)} - \bar{\alpha} \right)^2}{\frac{1}{n_k} + \frac{r_{\text{sum}}}{n}}}{\sum_{k=1}^m \frac{r^{(k)}}{1 - 1/n_k} \sum_{j=1}^{n_k} \left( \alpha_{k,j} - \bar{\alpha}^{(k)} \right)^2}$$

The distribution of  $V$  is very close to the F-distribution and by minimizing it, we can get a good estimation for the mean and variance of the individual physiological state values, therefore for similar parameters of the individual lag times.

The method was implemented in a user-friendly spreadsheet model (Baranyi and Pin, 1999) and tested on many datasets derived from laboratory experiments.

**10.** We developed formulae to obtain information also on the shape of the distribution of the individual lag times. The calculations were based on the situation that the initial cell number is random and follows the Poisson distribution, which can be achieved by serial dilutions (Baranyi et al, 2009). In this case, when the number of terms in a sum is also random, namely Poisson-distributed, then the moments of the sum of random variables can be calculated explicitly. We assumed that the distribution of the individual lag times could be described by a shifted gamma distribution and their mean value have the form  $L_g = T_{\text{shift}} + \beta\theta$  where  $\beta$  is the shape,  $\theta$  is the scale parameter of the distribution. Equating the theoretical and the sample moments, we obtained a method that was an extension, with more complex formulae, of the one described in the previous point. The procedure was validated by comparing results with those obtained by the maximum-likelihood method. The outcomes were similar but our method was superior in terms of speed and computational resources. This made it possible to build it, again, into spreadsheet applications, advantage of which is that in most food microbiology laboratories data are collected and stored in this way.

**Declaration**

The above results are my work. My co-authors were not mathematicians, but microbiologists. The exceptions are Carmen Pin, with whom we jointly worked on some formulae in point 7 (Baranyi and Pin, 2001); and Zoltán Kutalik, who developed the maximum likelihood program and carried out the comparison between the two methods discussed in point 10 (Baranyi et al, 2009).