

PREDICTIVE MODELLING TO CONTROL MICROBIAL HAZARDS IN THE FOOD PROCESSING INDUSTRY

József Baranyi

Institute of Food Research, Norwich, NR4 7UA, UK

Email: jozsef.baranyi@bbsrc.ac.uk

To be published in: Proceedings of Euroconference on Veterinary Public Health and Food Safety Assurance.

Summary:

The dramatic increase of applying mathematical concepts and computational techniques to food microbiology questions has led to a discipline called ‘predictive microbiology’. It is focussing mainly on the description of microbial responses to food environments by mathematical models. Its aim is more than the mere collection and computational representation of microbial observations, possibly mathematical interpolation. With the accumulation of data and experience, qualitative features are becoming constraints for the mathematical models to be created, thus moving towards mechanistic modelling. In this development, both mathematical models and microbiology databases have played crucial roles.

Keywords: predictive microbiology, mathematical model, microbial growth, primary model, secondary model

INTRODUCTION

Microbiology was a primarily descriptive science until the twentieth century. Mathematical techniques that today belong to predictive microbiology appeared in the 20’s, when the canning industry quantified the necessary heat treatment to eliminate *Clostridium botulinum* by log-kill units. The expression “12 log-kill” meant that the heating time should ensure that the concentration of the bacteria undergo a reduction of 12 decimal orders of magnitude. The heating time was calculated by assuming linear relation between log-concentration and time, if the heating temperature is constant. For some decades, thermal inactivation remained the only area where mathematical modelling played a significant role in microbiology.

A new era started from the work of Monod (1942), who described the relation between the bacterial biomass and the consumed substrate by a coupled differential equation. Monod’s work signalled a fast development of the application of mathematical techniques, first of all in biotechnology. In the 60’s, mathematicians and chemical engineers put the mathematical modelling of bacterial kinetics in a higher gear. A basic paper in *Mathematical Biosciences* (Frederickson, 1967) marked not only the start of a new journal but also induced dramatic increase in the needs of applying advanced mathematical modelling in microbiology.

Food microbiologists recognised the necessity of developing their own mathematical models and tools in the 80’s, when also the name “predictive microbiology” was coined. The name is rather unfortunate, since “prediction” is one of the main aims of any sort of modelling. Perhaps something like “Quantitative

Microbial Ecology of Food” would have been more adequate. However, the activity is more important than the name, and “predictive microbiology” gained a respectful status in the scientific literature (Fig.1).

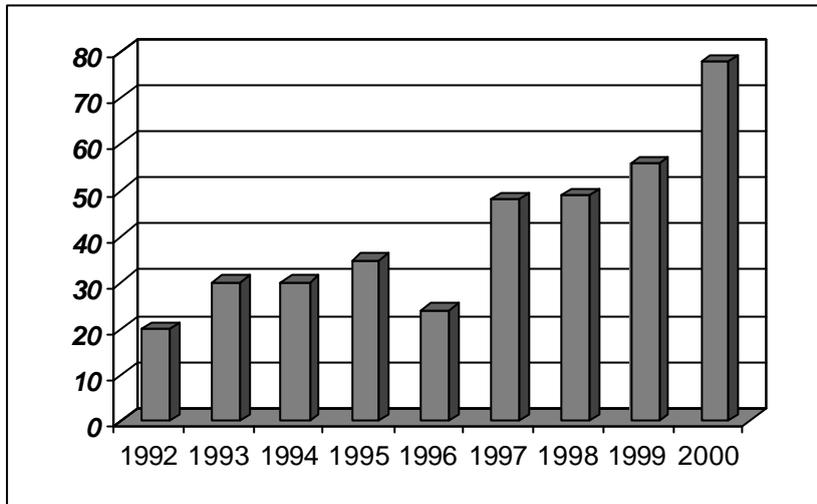


Fig.1. Number of papers with keywords “modelling” or “predictive microbiology” in the food science literature. (Source: Food Science and Technology Abstracts)

The main differences between mathematical modelling in biotechnology and food microbiology originate from the following points:

1. The studied ranges of cell concentrations differ in orders of magnitude. In biotechnology, it is commonly cell concentrations above 10^6 cells/ml, while it is mainly between 0 and 10^6 cells/ml in food microbiology. A consequence of this is that the models describe the real cell concentrations in the first, while their logarithm in the second case.
2. In biotechnology and fermentation technology, the aim is primarily to optimise certain product or production process; while in (predictive) food microbiology it is mainly to prevent bacterial growth.

In the latter case, the data and observations are more inaccurate, in a less controlled environment.

These differences were the main reasons why predictive microbiology could not take simply over mathematical models developed for biotechnology purposes but had to develop its own modelling techniques. Lots of these models use more empirical elements than models in biotechnology.

The increasing demand for mathematical modelling in the interest of microbiological safety was induced by the recognition that quality control of foods based on inspection of the final product was costly, laborious and inefficient. Mathematical models are quick and economical ways to objectively assess food safety. Their practical applications began to be materialised when powerful desktop computing had become commonly used. Lots of the models would have been impractical 30 years ago, because of the missing computing power to apply them.

PRIMARY MODELS

A mathematical model is developed by the way of mathematical abstraction. Because the real system is too complex, we must inevitably include simplifying idealizations and simplifications. These occur, for example, when one variable (index) quantifies a feature that is, in fact, a composite of several other aspects (for example, how to quantify the effect of food structure?), or when a variable is considered to be constant in time and/or homogeneous in space although it is known to be in fact time-dependent and heterogeneous. The extent of this simplification depends on theoretical and practical considerations such as the available mathematical techniques, computing power and available observations on the system.

Mathematical models are frequently classified as mechanistic and empirical models. Empirical models are only expected to accurately describe a set of observations, without taking into account the intrinsic mechanism by which these data are generated. A mechanistic model describes rather the process, either directly observable or unobservable, that generates those data. In practice, purely mechanistic models are rare, rather a mixture of the two is applied, possibly closer to one than to the other. Examples for models involving more empirical than mechanistic elements are those used in predictive microbiology, initially aiming at the pure collection and smoothed (i.e. “noiseless”) representation of computerised microbial data. However, with the increase of these data, more and more experience is accumulated and certain qualitative features become as “compulsory” for the models to be created. Such a feature, for example, is the existence of pure exponential growth in a constant environment supporting growth. A consequence of this assumption is that any mechanistic model should give linear relation between the logarithm of the cell concentration and time in an ideal situation. This is a starting point, which is to be modified in order to describe more and more complex situations as other considerations (lag, stationary phase, changing environment, etc) are also taken into account.

Frequently, model development is an iterative process going through a 'learning curve', when initial, empirical models (describing observations purely quantitatively) can help to define certain qualitative features of a more mechanistic model to be developed. It is also desirable to embed the model into more general principles of science and to make it open for further developments as the quantity and quality of information on the system increases.

A reasonable approach to build a predictive model on microbial growth must start with characterizing the classical growth curve (Pirt, 1975; see Fig.2). These models are frequently classified as primary models.

Typically, it is plotted on the log-scale, where the “logarithm of the cell concentration vs. time” gives a sigmoid shape. In what follows, we mean the log-transformation as the natural logarithm. We will denote the cell concentration by x and its logarithm by $y = \ln x$. Therefore, the slope of the curve at any point can be called “specific growth rate”, since

$$\frac{d(\ln x)}{dt} = \frac{dx/dt}{x}$$

The specific growth rate can be conceived as increase of cell concentration *per* unit time *per* cell number that produced that increase.

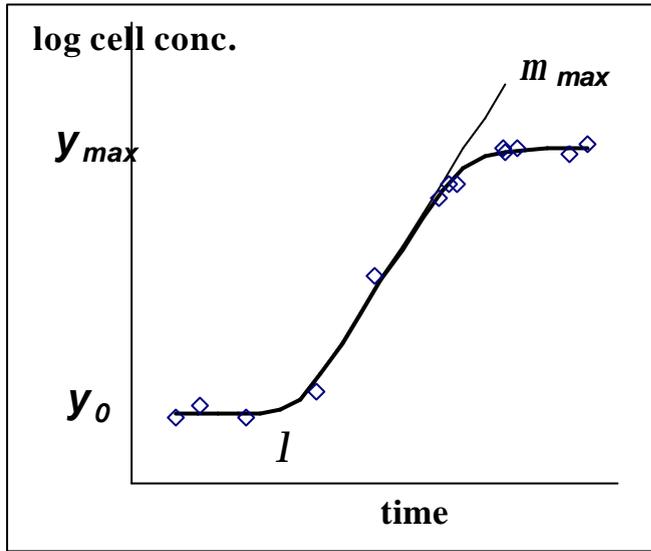


Fig.2. Classical bacterial growth curve of sigmoid shape, characterizable with four parameters.

A basic belief commonly accepted in bacterial growth modelling that there exists a pure exponential growth phase where the specific growth rate is constant. This is preceded and followed by lag phase and stationary phase, respectively, with lower specific rates. In what follows, we use the following notations (see Fig.2):

- y_0 : natural logarithm of the inoculum level;
- y_{max} : natural logarithm of the final cell concentration level in the stationary phase;
- m_{max} : The maximum specific growth rate measured in the exponential phase. It can be obtained by the slope of the tangent drawn to the inflexion point of the “ln x vs. time” sigmoid curve.
- l : The time marking the end of the lag period. It can be obtained as the intercept of the tangent, drawn to the inflexion point of the “ln x vs. time” sigmoid curve, and the ln-level of the initial concentration: $y_0 = \ln$ (inoculum level).

These four parameters can satisfactorily describe the basic bacterial growth curve of sigmoid shape. Various primary growth models were analysed, for example, by Zwietering *et al.* (1991).

From mechanistic point of view, the four parameters can be categorised as follows. The maximum specific growth rate is a so-called autonomous parameter (see Baranyi *et al.*, 1993), characterising purely the ability of the bacteria to grow in the given environment, independently of the history of the cells. This assumption reflects the belief that the cells sooner or later grow at a specific rate determined by the actual growth environment, after a possible adjustment to it.

The final cell concentration is also an autonomous (history-independent) parameter, but much less important, from application point of view, than the maximum specific growth rate. Namely, after a certain, high level of cell

concentration, the food is spoiled or has superseded the infective dose anyway and the refinement of that stage is not important anymore, for food microbiology. As has been mentioned, the interest in lower cell concentration is characteristic to predictive microbiology, as opposed to biotechnology.

The initial cell concentration is obviously purely history dependent. In fact, in experiments, it is set up by the experimenter, and can be relatively easily estimated. In real food, however, its estimation can be complicated, which can cause difficulties when estimating the error of predictions of bacterial concentration in the actual environment.

The most difficult parameter, from modelling point of view, is the lag parameter, because both the history and the actual environment affect it. This is demonstrated in Fig.3.

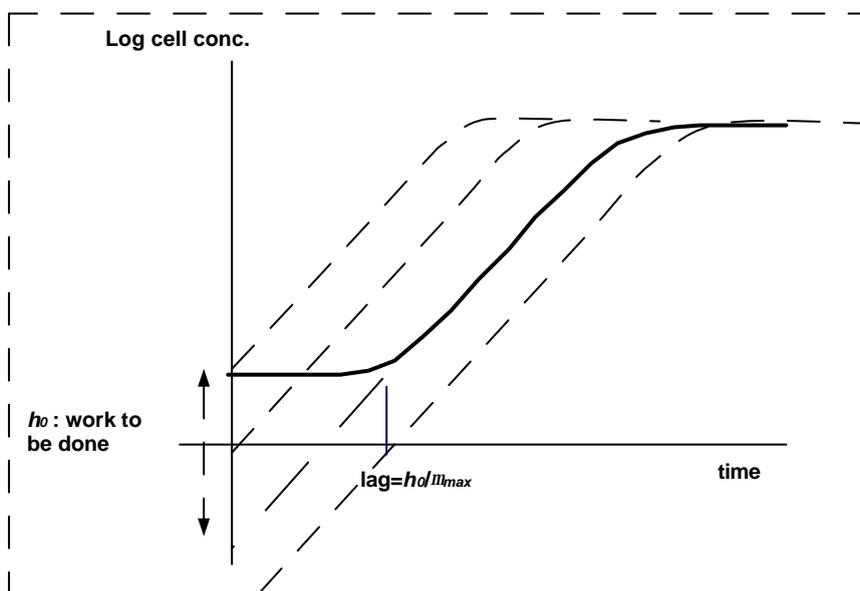


Fig.3. The actual growth curve adjusts to the field of potential growth curves which could have been obtained from different inoculum levels but with no lag.

To overcome this difficulty, Baranyi and Roberts (1995) re-parameterised the system and introduced the $h_0 = m_{max} \lambda$ quantity. If the lag and the maximum specific growth rate are inversely proportional (equivalently: if the relative lag defined as *lag/generation time of the exponential phase* is constant, which is observed by many researchers), then the parameter h_0 is constant and can quantify the work to be done during the lag phase. A rescaling of the h_0 parameter, $a_0 = \exp(-h_0)$ is a sort of “suitability” parameter, between 0 and 1, quantifying how much the history of the cells suitable to the actual environment. $a_0 = 1$ means optimum history, when there is no lag at all ($\lambda = 0$); and $a_0 = 0$ marks the infinitely long lag situation. Therefore, the system has two initial values: y_0 and a_0 . With this concept, the lag obviously depends on both history and the actual environment shown by the simple formula:

$$\lambda = \exp(-h_0) / m_{max}$$

It can be easily shown that the $a_0 = \exp(-m_{max} \Delta)$ quantity expresses the fraction of cells that would have been able to grow into the same curve without lag. Therefore, for example, $a_0 = 0.04$ means that if only 4% of the cells grow, they would reach a certain (high) concentration level at the same time as the actual growth curve, if those 4% can grow without lag.

The significance of this formulation is that it draws the attention to the history effect and the a_0 parameter (also called physiological state) is the first attempt to quantify that effect.

Primary inactivation or survival models can be treated similarly to the primary growth models. A thorough analysis on them is provided by Casolari (1988). The big difference between growth and death modelling is apparent when comparing the stationary phase of the growth curves and the “tailing” of the death curves, following the exponential phase in both cases. The maximum population level is an autonomous parameter of the growth models but this is not the case with the tailing of survival curves. Besides, at low cell concentrations, stochastic, rather than deterministic models should be used, as explained by Renshaw (1991).

SECONDARY MODELS

The microbial growth in food is affected by various factors: temperature, pH, water activity, atmosphere composition, preservatives, competition and inhibition of organisms (mainly via metabolites). The primary models describe the microbial behaviour under the assumption that the environmental factors are constant during the growth curve. As mentioned previously, a growth curve is identified with two history-dependent parameters (y_0 , the initial level of the natural logarithm of the cell concentration, and a_0 , characterising the suitability of the history of the cells to the actual environment); and two history-independent parameters (y_{max} , the final level of the natural logarithm of the cell concentration, and m_{max} , the maximum specific growth rate).

The history-dependent parameter y_0 are sometimes easy to estimate (for example, if a designed growth experiment is modelled), sometimes difficult (especially in practice, in real food). Generally, the main obstacle to practical modelling of the history-dependent parameters is that we don't know enough about the history of the cells. The autonomous parameters are easier to model in practice, for obvious reasons.

The models describing how the factors of an actual growth environment affect the growth parameters are frequently referred as secondary models. The structure of secondary models can be expressed by

$$g(p) = f(\underline{\mathbf{x}})$$

where

- p is a parameter of the primary model;
- g is a link function (used for numerical reasons only);
- $\underline{\mathbf{x}} = (x_1 \dots x_n)$ is the vector of environmental factors of the actual environment;
- f is a suitable multivariate function.

For the model parameter p , the maximum specific growth rate (or its reciprocal, the mean generation time) is studied most commonly, since that depends only on the actual environment.

The most frequent link functions are the square root and natural logarithm. Their use is justified by practical regression and numerical stability aspects; from mechanistic point of view, they don't have any significance.

The most important environmental factors are temperature, pH and water activity. Frequently, newer environmental factors are added only to modify a basic secondary model in such a way that the more-factor-model gives back the simpler one as a special case.

The f function is the core of the secondary model. For it, sometimes, a standard multivariate polynomial is used, whose advantage is that it allows carrying out linear regression during model fitting. Its disadvantage is that it is purely empirical and, as such, cannot be used outside the interpolation region defined by observed data. The construction of the interpolation region, in multivariate case (Fig.4.), can be a rather complex problem (Baranyi *et al.* 1996).

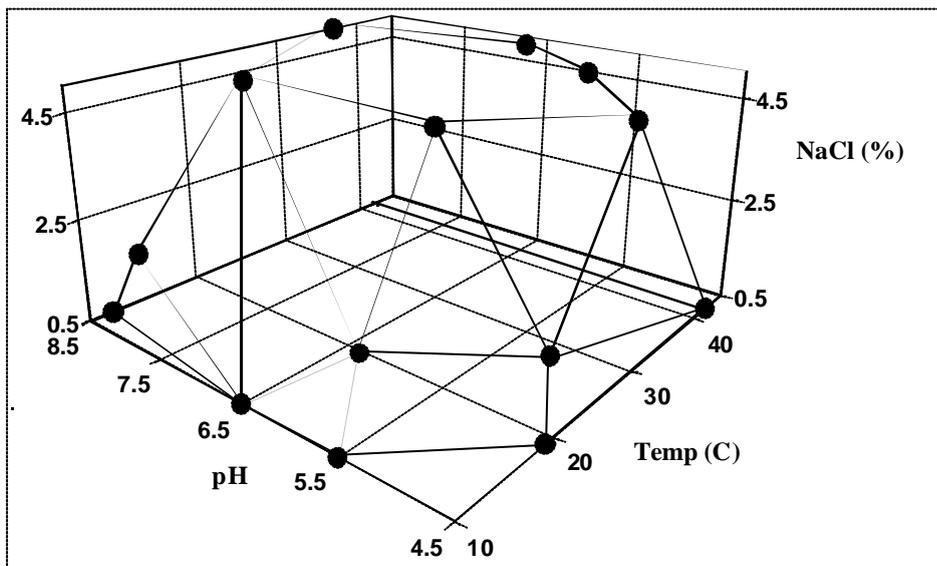


Fig.4. Strict interpolation region in the 3D space of Temp, pH and salt. It is the minimum convex polyhedron (convex hull) containing the locations of all the observations that provide the data for the regression.

McMeekin *et al* (1993) lists several possible secondary models. Two distinct approaches can be noticed in these models:

1. Models based on polynomials.

This approach considers only the computational easiness. Its main advantage is that fitting polynomial models leads to linear regression that can be readily solved by commonly available statistical packages. The disadvantage is that it's difficult to give biological meanings to the model parameters.

2. Models based on parameters with biological meanings.

This approach prefers those models whose parameters have biologically interpretable meanings. Such frequently used family of models are the so-called “cardinal-value-models” (Ratkowsky, 1983; Rosso, 1993), where the minimum, maximum, and optimum values of the environmental factors (which is, in the case of these authors, the temperature only) are among the model parameters. Fitting the model to data, generally, leads to non-linear regression, and the biological meaning makes it easier to get initial estimations of the parameters (which would not be necessary for linear regression). A typical method of creating a multivariate secondary model is multiplying one-factor models (Rosso *et al.*, 1995; Houtsma *et al.*, 1996.) The disadvantage of this, so-called “Gamma-concept” (Wijtzes *et al.*, 1995) is that it is not able to model the dependence of the cardinal values on other environmental variables.

CREATING PREDICTIVE SOFTWARE PACKAGES

Predictive software packages are based on two pillars: mathematical models and microbiological data. The process of creating a package can be summarised in the steps below, in order of complexity:

1. **Raw data.** Microbiological data collected by microbiologists, in various formats and media.
2. **Database.** A systematically structured database with rigorous syntax and semantics. Its creation needs understanding the data, expert judgement and computational skill.
3. **Browser.** A computer program to navigate in the database. This allows to search particular records satisfying certain query conditions and displaying the contents of those records in a user-friendly format (see Fig.5.)
4. **Predictive software.** Generating predictions using mathematical models. These are either interpolated values given by mathematical equations, or more sophisticated predictions (such as under dynamic conditions, probability estimations, growth/no growth boundary etc).

Since the 1980s, significant amount of data have been collected in internationally recognised research laboratories to support predictive microbiology and quantitative microbial risk assessment studies. Much of these data have been published or publicly available, but they are not necessarily accessible; the main obstacle being the lack of standard data format for recording these data.

Such database format has been developed at the Institute of Food Research (IFR, Norwich, UK). The format has been adopted by the UK Food Standards Agency for the data collected in their Predictive Microbiology Programme, 1985-1992. The USDA-ARS Eastern Regional Research Centre, Wyndmoor, PA, USA have also joined the initiative, and they transformed their raw data behind the Pathogen Modelling Program (PMP, see <http://www.arserrc.gov/mfs/pathogen.htm>) into this format. The result is a COMmon relational dataBASE called **ComBase**, which is now freely available on <http://wyndmoor.arserrc.gov/combase/>. Under the funding of the European Union (QLAM-2001-00513: e-ComBase), many EU institutions are also

joining *ComBase*. By the end of 2003, ComBase will contain up to 20,000 full bacterial growth and survival curves and some 8000 records containing growth/survival rate parameters. This unified international database helps to standardise the work of different risk assessors, with obvious positive implications on international food trade. For more information, see <http://www.ifr.ac.uk/combase/>.

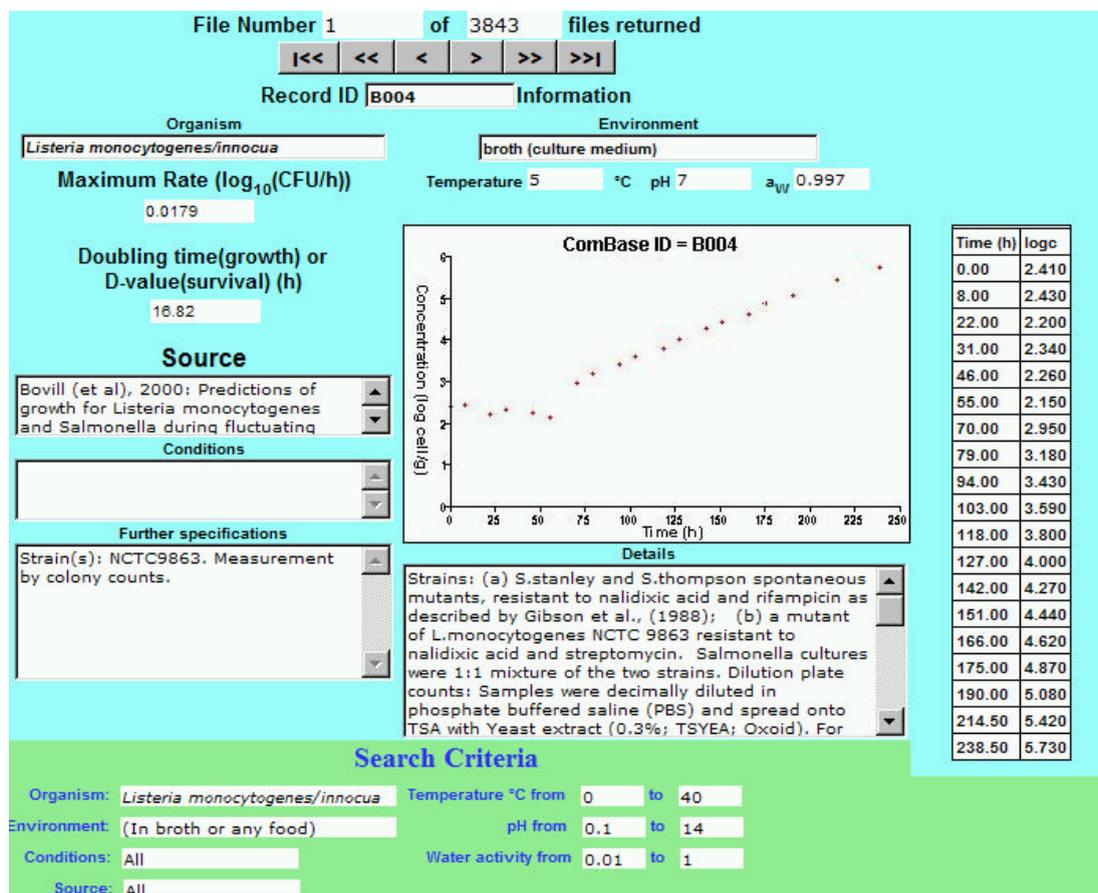


Fig.5. User-interface of the browser of the *ComBase* database on microbial responses to food environments.

MEASURING THE PERFORMANCE OF PREDICTIVE MODELS

An example for measuring the possible error generated by the primary model is shown in Fig.6. The Gompertz function, used by the Pathogen Modelling Program, tends to overestimate the maximum specific growth rate. As can be seen, the bias is about 0.1, in terms of log-doubling time. It can be easily calculated that this means, on average, about 20% relative difference (the ration between the difference and the average value of the predictions). This is why its predictions are higher than those produced by Growth Predictor (<http://www.ifr.ac.uk/Safety/GrowthPredictor/>), which is based on the model of Baranyi and Roberts (1995). Note that the comparison was made on the common interpolation region of temperature, pH and NaCl concentration.

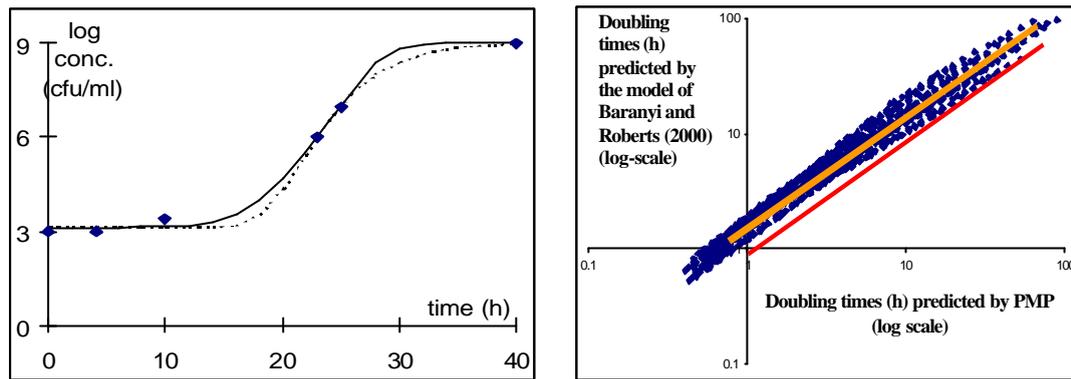


Fig.6. Gompertz' sigmoid curve (broken line in the first plot) tends to overestimate the exponential growth rate, which is especially apparent if there are only a few data in the exponential phase. This results in too cautious (conservative) predictions of the PMP software package.

Similar comparisons can be made between predictions based on data measured in broth and measured in real food. Baranyi and Pin (2000) estimated that, as an overall picture, the primary model is responsible for about 10% relative error, while each environmental variable of the secondary model accounts for another 10%. The difference between real food and broth is responsible for all the rest, which can be considerable, depending on the type of food.

As mentioned, only the specific growth/death rate (equivalently, the mean generation time / D-value), characterising the relation between the given bacterial population and the actual environment, can be predicted reliably. Even with that, the error margin can be rather big with more and more complex foods. However, predictive models can be used not only to predict the growth parameters in absolute sense, but also in relative sense. That means that the relative effect of the change of one environmental factor in food can be estimated by using models based on broth measurements. This is similar to the z -value concept in linear thermal inactivation models, where the question is, what temperature increase causes one order of magnitude decrease in the D -value. This question can be answered by the “generalised z -value” idea of Pin *et al* (2001), which can be used even if the predictions are (consistently) conservative.

CONCLUSIONS

What has been achieved

Predictive modelling has been established in food microbiology as a useful tool to decrease the cost and inaccuracy of controlling microbial safety and quality of food. Its main stakeholders are industry, government offices and authorities, as well as academia. With the dramatic development of desktop computing and Internet connection, the commonly available international databases and mathematical models are going to make predictive models become as everyday tools for researchers, authorities and students of food microbiology.

What has been neglected and what needs to be done

As Baranyi and Pin (2000) remarked, “*mathematical modelling is the art of omitting the unnecessary*”. However, “what is unnecessary” depends on the aim of the modelling, which can get more and more refined. On the other hand, the modelling tools depend on practicalities such as applicability, user-friendliness, speed, available resources, etc.

As said above, mathematical models are created by the way of abstraction, which contain simplifications, neglecting elements of the studied system that could have made the model impractically complex. Such intentional neglect ions are, typically:

1. Difference between laboratory experiments and real food situations, which can be much more complex.
2. The microbial environment is characterised only by a few factors/conditions (candidates for study: interactions of different organisms, communication-effect, food structure).
3. The history-effect is not taken into account satisfactorily.
4. Stochastic elements, such as variation of strains, variability of individual cells have not been characterised yet.

The points above readily give the tasks, too, what need to be done. Besides, it can be predicted that modern computational tools, such as Internet-based databases and integrated software packages (together, for example, with risk assessment and HACCP tools) will be among the future results. This will require interdisciplinary collaboration between food microbiologists and mathematicians, food technologists and statisticians, applying biomathematics and bioinformatics.

REFERENCE

- Baranyi J. and Roberts T.A. (1995). Mathematics of predictive food microbiology. *Int. J. Food Microbiol.* 26. 199-218.
- Baranyi J. and Roberts T.A. (2000). Principles and application of predictive modelling of the effects of preservation factors on micro-organisms. In: *The microbiology safety and quality of foods*. Ch. 18. Ed. Lund B., Baird-Parker and Gould, G. Aspen Publishers Inc.
- Baranyi J. and Pin C. (2000). Modelling microbiological safety. In: *Food Process Modelling*. Eds: L.M. Tijskens, M.L.A.T.M. Hertog. Woodhead Publishing, London.
- Baranyi J. , Tamplin M. (2002). ComBase: A Common Database on Microbial Responses to Food Environments. *J. Food Prot.* (In press).
- Casolari A. (1988). Microbial death. Chapter 7 in: *Physiological models in microbiology*. Eds: Bazin, M.J. and Prosser, J.I. Vol. II. CRC Peress, Boca Raton, Fla.
- Frederickson A.G., Ramkrishna D. and Tsuchiya H.M., 1967. Statistics and Dynamics of Procarvotic Cell Populations. *Math. Biosci.* 1. 327-374.
- Houtsma P.C., Muermans M.L., Rombouts F.M., and Zwietering M. (1996). Model for the combined effects of temperature, pH, and sodium lactate on growth rates of *Listeria innocua* in broth and Bologna-type sausages. *Appl.Env.Microbiol.* 62. 1616-1622.
- McMeekin T.A., Olley J.N., Ross T. and Ratkowsky D.A. (1993) *Predictive Microbiology*. John Wiley & Sons Ltd. Chichester, UK.

- Monod, J. (1942). Recherches sur la croissance des cultures bactériennes. Hermann. Paris.
- Ratkowsky D.A., Lowry R.K., McMeekin T.A., Stokes A.N., Chandler R.E., (1983). Model for bacterial culture growth rate throughout the entire biokinetic temperature range. *J. Bacteriol.* 154. 1222-1226.
- Renshaw, E. (1991). Modelling biological populations in space and time. Cambridge University Press. Cambridge.
- Rosso L., Lobry J.R. and Flandrois P. (1993). An unexpected correlation between cardinal temperatures of microbial growth highlighted by a new model. *J.Theor.Biol.* 164. 447-463.
- Rosso L., Lobry J.R., Bajard S. and Flandrois J.P. (1995). A convenient model to describe the combined effects of temperature and pH on microbial growth. *Appl.Env.Microbiol.* 61. 610-616.
- Tsuchiya H. M., Fredrickson, A. G. and Aris, R., (1966). Dynamics of microbial cell populations. In: *Advances in Chemical Engineering* 6. 125-206.
- Pin C., García de Fernando G., Ordoñez J.A. and Baranyi J. (2001). Applying a generalized z-value concept to quantify and compare the effect of environmental factors on the growth of *Listeria monocytogenes*. *Food Microbiology* 18, 539-545.
- Pirt S.J., (1975). Principles of microbe and cell cultivation. Blackwell. London.
- Wijtzes T., de Wit J.C., Huis in't Veld J.H.J., van't Riet K. and Zwietering M.H. (1995). Modelling the growth of *Lactobacillus curvatus* as a function of acidity and temperature. *Appl. Environ. Microbiol.* 61. 2533-2539.
- 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.