BINARY Vol. 3 © 1991 Academic Press Limited

Comparison of Estimates of Monod's Growth Model Parameters from the Same Data Set

J.R. Lobry and J.P. Flandrois[1]

Laboratoire de Biométrie, CNRS URA 243, Université de Claude Bernard, 43 Bd du 11 Novembre
69622 Villeurbanne Cedex, France
Electronic mail: DYNBACT@FRCISM51 (EARN)
[1]Faculté de Médecine Lyon-Sud, Laboratoire de Bactériologie, CNRS URA 243, Hôtel Dieu
69288 Lyon Cedex 02, France

Received: 28 June 1990; revised 29 October 1990

This report studies the estimation of Monod's bacterial growth model parameters μ_m and K_s using Monod's original data for growth of ***Escherichia coli*** *in a batch experiment with lactose as the limiting nutrient. The sum of squared residuals was used to determine whether there was a good fit between observed and theoretical points. Theoretical points were obtained by numerical integration of the model. The minimum of the sum of squared residuals was located with a nonlinear regression algorithm. The sum of squared residuals was systematically evaluated for different μ_m and K_s values to build a confidence region. This confidence region is wide, the value for μ_m is in the range 0.79—1.00 h^{-1}, and for K_s in the range 10—37 mg lactose l^{-1}. A comparison with previously published estimates from the same data shows that linearization methods should be avoided because they yield values outside the confidence region. It is argued that a systematic exploration of the sum of squared residuals is an efficient tool, especially for nonlinear models with few parameters.*

Introduction

Monod's bacterial growth model (Monod, 1941; 1949; 1950) defines the relation between specific growth rate and concentration of the limiting nutrient, and the relation between nutrient utilization rate and bacterial growth,

$$\begin{cases} \dfrac{dx}{dt} = x\mu_m \dfrac{s}{K_s + s} \\ \dfrac{ds}{dt} = - \dfrac{-dx}{Ydt} \end{cases} \qquad (1)$$

where x is the biomass concentration at time t, s the concentration of the growth limiting nutrient (or substrate) at time t, Y the yield coefficient, μ_m the maximum specific growth rate, and K_s the substrate concentration which supports half-maximum specific growth rate. A typical batch experiment starts with an inoculum ($x = x_0$) and an initial substrate concentration ($s = s_0$) at $t = 0$, and consists of observation of biomass and/or substrate concentrations until the final biomass is reached ($x = x_m$) when all the substrate is exhausted ($s = 0$). Units for biomass and substrate depend on the way growth is measured (Harris & Kell, 1985).

Applications of this model include biotechnological processes where various agents (antibiotics, enzymes, pesticides, proteins) are synthesized by micro-organisms, biological wastewater treatment processes (e.g. the activated sludge process) in which various compounds are removed from water by micro-organisms, and microbial ecology where nutrient-based models are used to make environmental management decisions.

In this paper methods used to estimate the parameters of Monod's model for batch culture data were investigated. To allow comparisons, original data from Monod, that were also used by Levenspiel (1980) and by Stiebitz *et al.* (1987), have been selected. The paper by Stiebitz *et al.* is especially interesting with respect to the number of different methods used. The method employed here differs somewhat from that used in the above-mentioned studies since we worked directly with biomass values instead of using an intermediate transformation to obtain μ versus s data.

Materials and methods

Data

The data set (Table 1) consisted of a batch experiment using *Escherichia coli* growing with lactose as the sole carbon and energy source. Initial lactose concentration was 200 mg l^{-1}. Note that time zero is not the true start of the experiment since Monod removed early points within the lag phase. One may ask why we did not use simulated data which included random error? Regarding the estimation of Monod's model parameters from batch culture data, this approach has already been taken at least three times before (Lederman *et al.*, 1976; Nihtilä & Virkkunen 1977; Holmberg, 1982). The problem is that conclusions vary from author to author because of differences in simulation conditions (parameter values, additive or multiplicative error models, intensity of variance of errors, size of time-sampling intervals). Since it is difficult to include realistic random errors, we have used a real data set, which has the additional advantage that it has been used before by different authors.

Data Processing

Computations were performed with an IBM RT/6150 computer with graphics drawn on a Macintosh IIcx computer. Programs were written in FORTRAN and are available on request, but it should be noted that subrou-

0226-304X/91/01020+04 $03.00/0

Table 1 Growth of *Escherichia coli* with lactose as limiting nutrient. x = biomass concentration, one unit of biomass corresponds to approximately 0.75 mg dry weight l^{-1} in Monod's optical density scale; s = concentration of growth limiting nutrient lactose. Data from Monod (1941), experiment number I in Table XIX p. 74 and Figure 18, p. 71, according to Levenspiel (1980) and Stiebitz *et al.* (1987)

Growth data		
Time (h)	x (biomass units)	s (mg l^{-1})
0.00	15.5	151
0.54	23.0	123
0.90	30.0	105
1.23	38.8	75
1.58	48.5	43.5
1.95	58.3	14.5
2.33	61.3	3.5
2.70	62.5	0.5

tine calls from IMSL 1.0 Library (Houston, USA) and specific instructions for interfacing with an AIX operating system were used.

Model Formulation

If the assumption is made that growth yield is constant, system (1) can be simplified as follows

$$\frac{dx}{dt} = x\mu_m \frac{(x_m - x)/Y}{K_s + (x_m - x)/Y} \qquad (2)$$

A more efficient form for integration is obtained since only one state variable, biomass, remains. The assumption that the yield coefficient is constant is well assessed only for energy-substrate limited growth (Wanner & Egli, 1990). Two requirements for using equation (2) are that: (i) the final biomass x_m is known, and (ii) the initial substrate concentration is known so that the yield parameter may be obtained from:

$$Y = \frac{x_m - x_0}{s_0} \qquad (3)$$

Integration

The differential equation (2) was numerically integrated using the IVPRK routine from IMSL which is based on a Runge-Kutta method (see Salmon, 1988). As a result we obtained a list of theoretical biomass values for a given set of parameter values.

Parameter Estimation

The usual Sum of Squared Residuals (SSR) was used to determine the goodness of fit. The SSR is the sum over all points of the squared difference between observed (x_i) and theoretical (x_i^*) biomass values,

$$SSR(\mu_m, K_s) = \sum_{i=1}^{n} (x_i - x_i^*)^2 \qquad (4)$$

The minimal value for the SSR was found using a modification (Brown & Dennis, 1972) of Marquardt's (1963) algorithm (routine UNLSF from IMSL). Values for x_0 and x_m have been fixed directly from data. It would be more rigorous to also consider them as parameters to be estimated. However, this increases computer-time without much improvement, since examination of the sensitivity functions (Holmberg, 1982) shows that the influence of remaining points on estimates of x_0 and x_m is of minor importance, which means, for instance, that a knowledge of the maximal biomass is not important for an estimation of initial biomass.

The value of the SSR was then systematically explored, setting parameter values by two overlapping loops. The aims were to check the minimum SSR value found previously and to delimit the confidence limits for parameter values. A 1-α confidence region for parameters is defined according to Beale (1960) by the set of parameter values so that the SSR is less than a threshold,

$$\left\{(\mu_m, K_s): SSR(\mu_m, K_s) \le SSR(\hat{\mu}_m, \hat{K}_s)\left(1 + \frac{p}{n-p} F^{\alpha}_{p;n-p}\right)\right\} \qquad (5)$$

where p is the number of estimated parameters, n the number of points in the data set, $\hat{\mu}_m$ and $\hat{K}_s$ the parameter values such that the SSR is minimum. Equation (5) gives a threshold value which indicates the level where the SSR surface has to be cut to obtain the contour line which delimits the confidence region.

Contouring programs typically encounter difficulties with long, thin segments of contours. The simplest way of improving resolution of the contour line is to increase the number of SSR evaluations in the region of interest, i.e. near the contour line (Figure 1). To do this, after a first grid computation of the SSR, each SSR evaluation was reconsidered to determine whether one of its four closest evaluations was on the opposite side of the threshold

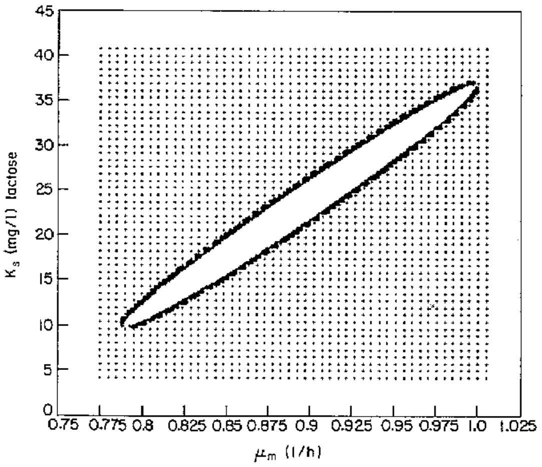

Figure 1 Systematic evaluation of the Sum of Squared Residuals (SSR) as a function of μ_m and K_s values. A cross (+) means that the SSR was found to be greater than the threshold of 3.2 as determined from equation (5). Corresponding pairs of values (μ_m, K_s) were then outside the confidence region. To increase the number of available points for interpolation of the contour line, the SSR was evaluated in its vicinity with improved resolution.

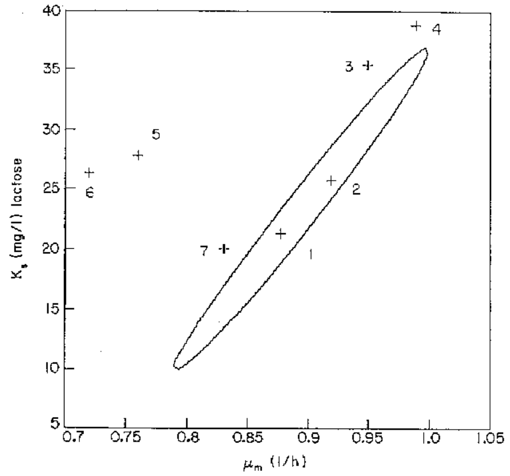

Figure 2 Estimates of Monod's growth model parameters μ_m and K_s from data in Table 1. The contour line encloses the 95% confidence region as determined in this paper. Crosses indicate the location of seven different estimates. 1: This work; 2—5: Stiebitz *et al.* (1987); 6: Levenspiel (1980); 7: Monod (1941). 2 and 6 are obtained from nonlinear regression from μ versus *s* data. 4—6 are obtained after linearization of μ versus *s* data according to 3: Eadie & Hofstee; 4: Langmuir & Hanes; 5: Lineweaver & Burk.

value. In the positive case, this means that the corresponding pair of values (μ_m, K_s) gives an SSR which is close to the contour line. It is then advisable to explore the SSR in its vicinity.

When the grid search was completed, a list of points was generated that were as close as possible to the contour line. To do this, we used interpolation between points on opposite sides of the threshold. With linear interpolation between two points, the position of the interpolated point is not accurate: from the quadratic behavior of the SSR surface, a linear interpolation generates points closer to the minimum than they should be, underestimating the area of the confidence region. A simple way to cope with this is to work with square-root values of the SSR for linear interpolation. In practice, when the grid is fine enough as in Figure 1, this correction is not important.

Results

The minimum SSR found by the nonlinear regression algorithm was at a level of 1.18 for μ_m = 0.878 h^{-1} and K_s = 21.2 mg lactose l^{-1}. The same value was obtained with systematic exploration of the SSR, which meant that convergence of the nonlinear regression algorithm was correct.

The critical level for the confidence region was then set at 3.2 according to equation (5) with α = 5% (Figure 2). The value for K_s was expected to be in the range of 10—37 mg lactose l^{-1} and the value for μ_m in the range 0.79—1.00 h^{-1}, equivalent to a doubling time in the range 42—53 min. As a matter of comparison, lactose concentration ranges from 4 to 100 µg l^{-1} in seawater (Mopper, 1980).

The confidence region was different from usual linear approximations of confidence regions since the location of the minimum was not at its geometrical center, because of the nonlinearity of the model. Parameters μ_m and K_s were positively correlated, meaning that an increase in μ_m may be partly compensated by an increase in K_s. This is a consequence of the model and should not be interpreted as having biological significance.

Estimates of parameters μ_m and K_s reported in the literature were never within this confidence region, except for estimate 2 from Stiebitz *et al.* (1987). Linearization methods gave poor results and even the graphical-empirical estimate from Monod was better. The fact that linearization methods should be avoided is clearly recognized (Dowd & Riggs, 1965; Bliss & James 1966; Tseng & Hsu, 1990), however they are still frequently used (Garfield, 1990).

It should be pointed out that reported estimates are not very distant from the confidence region: the main problem is that they have not been presented with confidence limits. Without confidence limits, the variability of results could be interpreted simply as an expression of microbial variability. However, in the present paper, this cannot be the case since we have used the same data set. The ratio of the maximum estimate to the minimum estimate was 1:4 for μ_m, and 1:9 for K_s. Differences in the parameter estimation method may therefore explain the variability of reported results, but only partially: for glucose-limited growth of *E. coli*, important variations in K_s values have been reported, from 8 mg l^{-1} to 0.2 mg l^{-1} (Koch & Wang, 1982).

Conclusion

The systematic exploration of the Sum of Squared Residuals is generally not held in high esteem by purists. This is because the amount of computing effort and the amount of storage required increases as a power function of the number of parameters involved. Furthermore, such a rough approach lacks elegance. These reasons are less valid when the model has few parameters. Moreover, with nonlinear models, this is the only way to obtain a 'true' confidence region for parameters.

The method used here to obtain a contour line of the confidence region is certainly not optimal. For instance, recursive magnification of the region of interest would be more efficient in terms of computer time and storage, but it is more difficult to program with non-recursive oriented language such as FORTRAN. The important point to remember is that we need to focus on the contour line because the confidence region is sometimes very narrow, especially when parameters are highly correlated.

Advantages of a systematic grid search are that it is easy to program, and more important, easy to understand. Determination of the precision of parameter values is straightforward. Over-parameterization of models is easy to detect: the confidence region extends to infinity when there are not enough data to estimate parameters.

Acknowledgements

This work was supported in part by a grant from API-SYSTEM Biomérieux, to J.R. Lobry. We wish to thank the editorial board and referees for helpful suggestions in drafting the manuscript.

References

BEALE, E.M.L. (1960). Confidence regions in non-linear estimation. *Journal of the Royal Statistical Society* **B 22**, 41-88.

BINARY Vol. 3 © 1991 Academic Press Limited

BLISS, C.I. & JAMES, A.T. (1966). Fitting the rectangular hyperbola. *Biometrics* **22**, 573-602.

BROWN, K.M. & DENNIS, J.E. (1972). Derivative free analogues of the Levenberg-Marquardt and Gauss algorithms for nonlinear least squares approximation. *Numerische Mathematik* **18**, 289-297.

DOWD, J.E. & RIGGS, D.S. (1965). A comparison of estimates of Michaelis-Menten kinetic constants from various linear transformations. *Journal of Biological Chemistry* **240**, 863-869.

GARFIELD, E. (1990). The most-cited papers of all time. *Current Contents/Life Sciences* **33**, 3-15.

HARRIS, C.M. & KELL, D.B. (1985). The estimation of microbial biomass. *Biosensors* **1**, 17-84.

HOLMBERG, A. (1982). On the practical identifiability of microbial growth models incorporating Michaelis-Menten type non-linearities. *Mathematical Biosciences* **62**, 23-43.

KOCH, A.L. & WANG, C.H. (1982). How close to the theoretical diffusion limit do bacterial uptake systems function? *Archives of Microbiology* **131**, 36-42.

LEDERMAN, T.C., HORNBERGER, G.M. & KELLY, M.G. (1976). The calibration of a phytoplankton growth model using batch culture data. *Water Air and Soil Pollution* **5**, 431-442.

LEVENSPIEL, O. (1980). The Monod equation: a revisit and a generalization to product inhibition situations. *Biotechnology and Bioengineering* **22**, 1671-1687.

MARQUARDT, D.W. (1963). An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial and Applied Mathematics* **11**, 431-441.

MONOD, J. (1941). Recherches sur la croissance des cultures bactériennes. Thèse de docteur ès sciences naturelles, Paris.

MONOD, J. (1949). The growth of bacterial cultures. *Annual Review of Microbiology* **3**, 371-394.

MONOD, J. (1950). La technique de la culture continue: théorie et applications. *Annales de l'Institut Pasteur* **79**, 390-410.

MOPPER, K. (1980). The monosaccharide spectra of natural waters. *Marine Chemistry* **10**, 55-65.

NIHTILÄ, M. & VIRKKUNEN, J. (1977). Practical identifiability of growth and substrate consumption models. *Biotechnology and Bioengineering* **19**, 1831-1850.

SALMON, I. (1988). Solving dynamic equations on a micro-computer. *Binary* **13**, 20-24.

STIEBITZ, O., WOLF, K.H. & KLÖDEN, W. (1987). Auswirkungen der Methode der Linearisierung des μ-Cs-Verlaufs auf die Parameteranpassung bei dem Monodschen Modell. *Acta Biotechnologica* **7**, 167-172.

TSENG, S. & HSU, J.P. (1990). A comparison of the parameter estimating procedures for the Michaelis-Menten model. *Journal of Theoretical Biology* **145**, 457-464.

WANNER, U. & EGLI, T. (1990). Dynamics of microbial growth and cell composition in batch culture. *FEMS Microbiology Reviews* **75**, 19-44.