UCRL-17953
cy. 2

University of California

Ernest O. Lawrence Radiation Laboratory

MATHEMATICAL REPRESENTATION OF BATCH CULTURE DATA

Victor H. Edwards and Charles R. Wilke

November 1967

RECEIVED
LAWRENCE
RADIATION LABORATORY

LIBRARY AND
DOCUMENTS SECTION

TWO-WEEK LOAN COPY

This is a Library Circulating Copy
which may be borrowed for two weeks.
For a personal retention copy, call
Tech. Info. Division, Ext. 5545

UCRL-17953
cy. 2

DISCLAIMER

This document was prepared as an account of work sponsored by the United States Government. While this document is believed to contain correct information, neither the United States Government nor any agency thereof, nor the Regents of the University of California, nor any of their employees, makes any warranty, express or implied, or assumes any legal responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by its trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or the Regents of the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof or the Regents of the University of California.

Submitted to Biotechnology and
Bioengineering

UCRL-17953
Preprint

UNIVERSITY OF CALIFORNIA

Lawrence Radiation Laboratory
Berkeley, California

AEC Contract No. W-7405-eng-48

MATHEMATICAL REPRESENTATION OF BATCH CULTURE DATA

Victor H. Edwards and Charles R. Wilke

November 1967

"Mathematical Representation of Batch Culture Data"

Victor H. Edwards*

and

Charles R. Wilke

Department of Chemical Engineering and
Lawrence Radiation Laboratory

University of California

Berkeley, California

ABSTRACT

A generalized logistic equation is proposed for the mathematical representation of batch culture kinetic data. Properties of the equation are discussed. A computer program is used to fit the generalized equation to both artificial and actual batch culture data. The equation is shown to be capable of fitting data exhibiting lag, exponential, deceleration, stationary, and death phases, as well as diauxic growth. The fitted equation is useful for differentiation, interpolation, and other manipulations of the data, and it is a convenient means of data storage.

* Present address: Department of Chemical Engineering
Cornell University, Ithaca, N. Y.

INTRODUCTION

The interpretation of batch culture kinetic data is a problem that arises frequently in many biochemical engineering studies ranging from basic research on microbial kinetics to work on the development of new processes or the improvement of existing ones. The critical kinetic parameters that must be extracted from these studies are yields and rates, because the economics of a process depend upon the amount of product that can be obtained from a given starting material and on the rate at which it can be produced. In many cases, a large number of batch experiments are performed to test the effects of different variables such as temperature and the concentration of various components of the growth medium. These results must then be differentiated graphically or by some other means if accurate values of growth, product formation, and substrate consumption rates are desired. Graphical differentiation of large amounts of kinetic data is tedious and susceptible to subjective errors. The use of interpolation formulae for numerical differentiation can lead to large errors if there is much scatter in the data or if the data are widely spaced.

These difficulties may be avoided by fitting a suitable equation to the kinetic data and differentiating the fitted equation to determine rates. By using a systematic fitting procedure such as the method of least squares, the influence of random errors in the data is greatly reduced and errors of judgment inherent in graphical differentiation of data are eliminated. With such an approach, more reliable data values can be weighted accordingly. When the magnitude of the errors

in the data is known, the quality can be directly tested. Finally, the process of data analysis can be made much easier and quicker by using a computer to perform the calculations. This paper reports the selection of a suitable equation for the representation of batch culture kinetic data, describes how it is used to analyze data, and discusses some typical results obtained in the application to both artificial data and actual data.

THEORY

The choice of the equation to be used to represent a set of data is crucial to the success of the undertaking. For example, one would not choose a periodic or discontinuous function to represent the change of sugar concentration with time in a normal batch culture of yeast. To simplify data analysis, it is desirable to use a single type of equation to represent all, or at least most, types of data encountered. The equation must be able to represent all phases of batch culture, including lag, acceleration, exponential, deceleration, stationary, and declining growth phases, as well as special phenomena such as diauxie. The need is thus for a flexible equation. Flexibility is also needed to ensure that the equation can be fit to the data with a small number of parameters without imparting a serious bias to the results. Once fitted to the data, the equation must be easy to use. These and some other desirable properties of an equation used for the analysis of batch culture kinetic data are summarized in table I.

~~Lets~~ Consider some particular examples. The simple polynomial shown in equation (1) is often used to represent data of unknown functional form.

$$y = a_o + a_1 t + a_2 t^2 + \ldots + a_n t^n \qquad (1)$$

where y = dependent variable

t = time of incubation

a_o, etc. = constants

The polynomial has many advantages, including ease of fitting and use, flexibility, and differentiability. However, it does not satisfy criteria 1, 3, and 5, because it cannot exhibit a stationary phase or a zero derivative at large times and a large number of terms may be required to obtain an adequate fit to typical batch culture data. Also, the fitted parameters have no direct physical meaning.

A functional form that would contain parameters having direct physical meaning can be obtained by assuming that the cell growth rate depends on the concentration of limiting substrate according to the familiar Monod equation.[1]

$$\mu = \frac{\mu_m S}{K_s + S} \qquad (2)$$

where μ = Specific growth rate

μ_m = Maximum specific growth rate

K_s = Monod constant

S = Concentration of limiting substrate

By assuming the yield of cells to be directly proportional to the amount of limiting substrate consumed, one can derive the following formula.[2]

$$t = \frac{(S_o + K_s)}{\mu_{max} S_o} \ln\left[\frac{X}{X_o}\right] - K_s \ln\left[\frac{X/Y - S_o}{Xo/_Y - S_o}\right] \qquad (3)$$

where S_o = initial substrate concentration

X = cell concentration at time T

X_o = initial cell concentration

Y = yield coefficient

The fittable parameters (K_s, μ_{max}, and Y) now have direct physical meaning, and from this standpoint, equation (3) is superior to the simple polynomial, at least when describing the time dependence of the cell concentration. Knowles, Downing, and Barrett have successfully applied this approach to the determination of the kinetic parameters for nitrifying bacteria in mixed culture.[3] In their work, a computer program was used to calculate the fittable parameters (K_s, μ_{max}, and Y) by a least squares technique.[3]

However, use of the Monod equation suffers from certain disadvantages. First, the equation used to describe the concentration of the limiting substrate is not of the same mathematical form as the equation used to describe the cell concentration so that similar but separate procedures must be employed to fit the two equations to the substrate and cell concentration data. In fact, the determination of the fittable parameters should be done for both substrate and cell concentration data simultaneously, fitting the data to both equations at once, so that only one set of values for K_s, Y, and μ_m will be obtained for a given batch culture experiment. The procedure of computing the parameters from the data is thus relatively complicated and may also involve the use of large amounts of computer time, particularly if the program is not designed carefully.[4]

The inconvenience associated with using Equation (3) to calculate cell concentrations or growth rates once the parameters have been determined is an additional drawback. However, the major disadvantage of the integrated Monod equation is its lack of flexibility. It cannot be used to represent either a lag phase or a phase of decline, and there are not enough fittable parameters to permit the description of systems not obeying Monod's equation or exhibiting a constant yield coefficient. Thus the method pioneered by Knowles, Downing, and Barrett, while useful for many cultures, is not sufficiently general for many other types of batch culture data.

A third type of equation that is often used to model population growth is the logistic equation. Among the first to use this equation with populations were Verhulst, Pearl and Reed.[5,6] It has the following form:

$$y = \frac{K}{1 + \exp(a_o + a_1 t)} \tag{4}$$

where y = dependent variable
t = time
a_o, a_1, K = constants

The logistic equation gives a sigmoidal curve when a_1 is negative, thus approximating a portion of a simple batch culture population. However, it does not allow for a lag phase or a phase of decline. Also, it is not as flexible as would be desired. But at least the parameters are capable of some physical interpretation. For example, if used to describe the cell concentration, the parameter K represents the value of the cell concentration in the stationary phase, the parameter a_1 represents the maximum specific growth rate, and the

third parameter, a_o, is related to the ratio of initial to final cell concentration. The logistic equation also offers the advantage of being comparatively easy to use, making it possible to calculate explicitly the values of the dependent variable and its derivatives.

The only serious drawback of the logistic equation is thus its lack of flexibility. This problem may be overcome by generalizing the equation through the use of a more general function for the argument of the exponential term in the denominator. Upon generalization, the logistic equation thus becomes

$$y = \frac{K}{1 + \exp\left[F(t_1)\right]} \tag{5}$$

where $F(t_1)$ = some function of time

$t_1 = t - t_L$

t_L = length of lag phase

$t > t_L$

Introduction of the new time scale, t_1, permits better modeling of the lag phase. An equation of this form retains a number of advantages for the purpose at hand, even for fairly complex $F(t_1)$. These advantages include ease of use, physical meaning for at least some parameters, and constant limiting values of y at small and large values of t_1.

Although better choices of F(t) may exist, a simple polynomial was selected, giving the following equation:

$$F(t) = a_o + a_1 t + a_2 t^2 + \dots a_n t^n \tag{6}$$

As was pointed out by Dewitt, this particular form of the generalized logistic equation can fit many types of data of sigmoidal form.[7] It is also easy to make initial estimates of the values of the parameters, simplifying its use. An additional advantage of this equation for modeling kinetic data is that at low values of y/K, exponential growth is well

approximated, gradually going over to product-inhibited growth at large time values. Equation (7) shows the dependence of the specific growth rate on values of y and t_1.

$$\frac{1}{y}\frac{dy}{dt} = -(1 - y/K)(+ a_1 + 2a_2t_1 + 3a_3t_1^2 + \ldots + na_nt_1^{n-1}) \tag{7}$$

The generalized logistic equation thus meets the need for flexibility, being especially adaptable to systems with product inhibition. However, it should be emphasized that only K, a_o, and a_1 have any direct physical (biological)meaning.

Properties of the generalized logistic equation were discussed by Dewitt.[7] An outline of the properties will be given here.

Equation (7) can be rearranged to give the following form

$$\frac{dy}{dt} = -y(1 - y/K)\, F'(t_1) \tag{8}$$

Because y is always greater than zero and less than K, the sign of the first derivative, dy/dt, is determined by $F'(t_1)$, which may take on successively positive and negative values at various times values. Points of inflection of Eq. (5) occur when Eq. (9) is satisfied.

$$y = K/2\ \left(1 - \frac{F''(t_1)}{(F'(t_1))^2}\right) \tag{9}$$

Depending on the number of terms in $F'(t_1)$ and their signs, a number of relative maxima and minima may be obtained. For example, for $F(t_1)$ a fifth degree polynomial, $F(t_1) = a_o + a_1t_1 + a_2t_1^2 + a_3t_1^3 + a_4t_1^4 + a_5t_1^5$, and up to 3 relative minima and 3 relative maxima may be obtained, as shown in Fig. 1. In Fig. 1, at low values of the independent variable, lower order terms predominate, with higher order terms becoming the determining factors at high values of the independent variables. For the case shown, a_1, a_3, and a_5 have negative values while a_o, a_2, and

a_4 have positive values. Of course, the values of a_o, a_1, ... a_n can take on any values desired to obtain curves of the desired type. The curve just shown demonstrates the maximum degree of flexibility obtainable with fifth order polynomial. In general, the maximum number of relative minima and maxima in the first quadrant may be at most one greater than the degree of the polynomial. The highest order term ultimately determines the asymptotes of the curve. Thus there are four possible results because a_n may be positive or negative and n may be odd or even. Figure 2 shows the four possible cases. Note that all values of the dependent variable lie between y = K and y = 0. For the modeling of bacterial growth, n was always given an odd value, values of 1, 3, and 5 being chosen, so that only curves of the second and the fourth types shown in Fig. 2 were used. Because the microbial growth is limited to the first quadrant, these two cases proved adequate to give the two desired alternatives of an asymptotic approach to a constant at large values of time and the case of an asymptotic approach to a zero value at large values of time. The first case could represent product concentrations versus time, while the second case could represent viable cell count versus time, which would approach zero concentration at very long incubation times. The flexibility of the model thus appears to be adequate to fit all types of batch culture data including data from complex fermentations. However, the non-linear least squares technique does not always succeed in fitting complex data, as will be discussed shortly, and better results could be obtained by employing a more sophisticated method of determining the parameters.

It should be recognized that the curves drawn in Figs. 1 and 2 do not represent actual plots of the generalized logistic equation.

They are hand-drawn curves, meant to illustrate the qualitative behavior of the equation.

METHODS

To reduce the labor of fitting the equation to the data, a computer program in Chippewa Fortran was written for use with a Control Data 6600 Computer. A method of non-linear least squares curve-fitting was used to fit the equation to the data. The method is that given by Scarborough. A linear least squares method will not work because the generalized logistic equation is nonlinear in the parameters to be computed. The nonlinear method consists of making initial guesses for the values of the coefficients and correcting these initial guesses by repeated iteration until a desired degree of convergence is obtained. The technique of making initial guesses is described in Appendix I. Corrections are obtained by assuming a Taylor series expansion of the fitting function in its coefficients, analogous to the Newton-Raphson technique of seeking roots of a nonlinear equation. The method and a detailed outline of the computer program is given by Edwards.[9] A brief description of the program is presented in Appendix II. Many groups of data may be analyzed in sequence by the program. The program fits succeedingly complex forms of the generalized logistic equation to the data, statistically testing each succeeding equation until a fit is obtained that satisfies a Chi-square test. The computer program also tabulates observed and calculated values of the data, tabulates values of the calculated parameters, and prepares graphs comparing calculated and observed data.

Because standard deviations were not available for much of the data discussed in this paper, estimates of the goodness of fit were obtained by calculating an average relative error.

The following formula was used:

$$R = 100\% \times \frac{\sum_{\text{All data points}} \left| \frac{\text{Calculated data value} - \text{Observed value}}{\text{Observed data value}} \right|}{\text{Number of data points}} \qquad (10)$$

where R = average relative error (%)

RESULTS

Both artificial and real batch culture kinetic data were supplied to the computer program to test its performance. Some of the results will now be described. Further applications are reported elsewhere.[9]

One test of the equation was made using data from an imaginary culture showing an exponential increase of cell concentration followed by a phase of constant cell concentration. For convenience, these data will be referred to as the exponential model.

The artificial data values are listed in the second column of Table II and are shown as open circles in Fig. 3. The exponential data were taken from a table of exponentials.[10] By plotting the imaginary data on semi-log graph paper, the data representing the exponential growth period fall on one straight line of positive slope and the points representing the stationary period fall on a second, intersecting horizontal straight line. The slope discontinuity prevented the continuous fuction from fitting the curve perfectly and provided a stress on the model. These hypothetical results would never be achieved in practice, but a close approximation might be made by a culture growing on a synthetic medium limited by a carbon source if the inoculum were a young, adapted, and exponentially growing culture that would continue growing exponentially

upon inoculation. Cell mass versus time would most likely follow a course well approximated by an exponential and a resting phase, although cell number might not, there being a tendency for cell division to continue for a short period even after exhaustion of the carbon source. Figure 3 and Table II compare the exponential model to the curves fit by the computer for the cases of first degree polynomial (which is equivalent to the simple logistic equation) with a fifth degree polynomial in the generalized form of the equation. The first degree equation is satisfactory in the low values of cell concentration, but it is inadequate at high cell concentrations, giving an average relative error of 6.9%. In the case of the fifth degree polynomial, the fit is good throughout, with an average relative error of less than 2%. The equations fit to the data are, respectively,

$$N = \frac{239.966}{1 + \exp(5.5530-0.5389t)} \tag{11}$$

$$N = \frac{151.473}{1+\exp(5.01945-.60037t+.10757t^2-.040219t^3+.0059485t^4-.00031375t^5)} \tag{12}$$

Use of Eq. (12) is clearly inconvenient without use of a computer or desk calculator. The fit of the specific growth rates calculated by computer are compared to the exponential model in Fig. 4 and Table III. As can be seen, the first degree equation again fails, particularly at high values of time and cell concentration, but the fifth degree equation does a good job of representing a general form of the variation of the specific growth rate with time. Average relative errors in calculated specific growth rates were 9.6% and 5.4% for the first and fifth degree cases; better results could be obtained by smoothing of the specific growth rate data obtained by computer.

A second test was made using the batch culture kinetic data of Luedeking.[11] Luedeking studied the lactic acid producing bacteria, Lactobacillus delbruckii. His data were especially useful as a test because he also tabulated graphically-measured growth rates.

The generalized logistic equation fit to the cell concentration data of Luedeking's run number 13 has the following form:

$$N = \frac{9.4906}{1+\exp(4.6029-.4622t+.10374t^2-.03135t^3+.003279t^4-.1.192\times10^{-4}t^5)} \qquad (13)$$

The actual data are compared with the values calculated with equation (13) in Fig. 5 and Table IV. An average relative error of 4.1% resulted with very good agreement between calculated and observed data over the entire range of values. Growth rates determined from the data by graphical differentiation by Luedeking are compared with values calculated by analytical differentiation of Eq. (13) are compared in Fig. 6 and Table IV. Once again, very good agreement is obtained confirming the adequacy of the generalized logistic equation.

Figures 7 and 8 show the results obtained by fitting the generalized logistic equation to some data obtained by Edwards with cultures of a salt-tolerant strain of sulfate-reducing bacteria.[9] Figure 7 is a graph of the cell concentration (as measured by an optical counting technique) as a function of incubation time. Data are shown for cultures at three different initial pH values. A third-degree generalized logistic equation was fitted to the data for each culture. The resulting equations are given in Fig. 7; the curves represent the values calculated with each equation. Two important points should be mentioned in connection with these data. First, the data were not taken as frequently as would have been desirable and they are somewhat scattered due to the limitations of

the optical counting technique. Second, the data point taken after 3.9 days for the culture with an initial pH of 8.60 seems to be a bad data point, appearing to be low by a factor of about two. The datum was included to see how it would affect the results. In spite of these two weaknesses in the data, the generalized logistic equation fitted the data very well, giving Chi-square values with probabilities of >98%, >95%, and >80% for the experiments at initial pH values of 7.60, 7.99, and 8.60, respectively. The lower probability obtained for the fit to the culture with an initial pH value of 8.60 resulted from inclusion of the questionable data value observed at 3.9 days. There is further evidence for the success of fit. Qualitative examination of the data in Fig. 7 suggests that the only major effect of the initial pH in this experiment was to alter the length of the lag phase. Otherwise, rates and yields appear to be about the same for all three cultures. Specific growth rates calculated by the computer from the differentiated form of the three fitted equations are plotted in Fig. 8 as a function of incubation time. It can be seen that the calculated specific growth rates confirm quantitatively what visual examination of the data suggests. The magnitude and shape of the three curves in Fig. 8 are all very similar, except that they are somewhat displaced in time. Again the effect of including the questionable point (for the culture with an initial pH value of 8.60) is apparant from the shape of the specific growth rate curve calculated for that culture at low values of incubation time and from the broadening of the curve. However, it is encouraging that inclusion of the point does not appear to influence the values of the specific growth rate calculated for higher incubation times.

Figure 9 illustrates the application of the generalized logistic

equation to product concentration in batch cultures. The data are from a different set of experiments with the same strain of sulfate-reducing bacteria used in the work just described.[9] In this case, the initial concentration was varied and other variables were held constant. As before, data were not spaced close enough to permit very accurate fitting, yet good results were still obtained using the generalized logistic equation. The three curves drawn in Fig. 9 represent the following three equations:

For 0.1% yeast extract,

$$\left[S^{--}\right] = \frac{24.01\text{mM}}{1 + \exp\left[6.4538 - 2.3731t + 0.2402t^2 - 0.009275t^3\right]} \tag{14}$$

For 0.5% yeast extract,

$$\left[S^{--}\right] = \frac{28.33\text{mM}}{1 + \exp\left[4.5557 - 1.0869t_1 + 0.01523t_1^2 - 7.7657\times10^{-5}t_1^3\right]} \tag{15}$$

$$t_1 = t - 1.1 \text{ days}$$

For 1.0% yeast extract,

$$\left[S^{--}\right] = \frac{31.55\text{mM}}{1 + \exp\left[21.998 - 10.3017t + 1.9084t^2 - 0.13018t^3\right]} \tag{16}$$

Some slight distortion can be seen between six and eight days for the culture having an initial yeast extract concentration of 0.5% and between seven and nine days for the culture with an initial yeast extract concentration of 1.0%. The distortion, though slight, was reflected in Chi-square values of much lower probability for equations (15) and (16). However, the results were quite good in view of the limited number of data points available to be fitted: they illustrate the usefulness of the generalized logistic equation in extracting information from limited experimental data.

All tests up to this point were made using data that approached a

positive, non-zero constant at large values of time. However, as was pointed out in the section on the mathematical properties, the generalized logistic equation can also approach zero at large times. This property is important because variables such as viable count tend to fall off to zero values in the period after the stationary phase. Also, intermediate products can accumulate in the early phases of a culture and then be consumed before the fermentation is complete. Thus, mathematical representation of the concentrations of such intermediate products requires a function capable of showing a maximum and then decreasing to zero or some other low value. An example is the diphasic bioxidation of glucose to 5-ketogluconic acid by Acetobacter suboxydans, which shows an accumulation of gluconic acid.

Synthetic data were assumed with a lag phase of 10.5 time units and having a bell-shaped curve. A generalized logistic equation was then fitted to the data. Table V and Fig10 compare the test data with the fitted equation. The fitted equation has the form

$$y = \frac{1.3805 \times 10^8}{1 + \exp(4.5593 - .86477t_1 - .64989t_1^2 + .1008t_1^3)} \qquad (17)$$

$$t_1 = t - 10.5 \text{ hr}$$

Good agreement between the test data and Eq.(17) is evident.

Thus the bell-shaped curve may be added to the list of curves that the model is capable of fitting.

As mentioned in the discussion of the properties of the generalized logistic equation, it has the capability of fitting complex curves. Hosler and Johnson obtained fairly complex data in their study of the penecillin fermentation.[12] In their work, mycelial nitrogen showed diauxie at intermediate incubation times and decreased at long incubation

times. The computer program fit a third degree model to the data with the following form:

$$y = \frac{2369.4}{1+\exp(3.1794-.05681t-3.37\times10^{-4}t^2 + 4.66\times10^{-6}t^3)} \tag{18}$$

Table VI and Fig.11 compare observed and fitted data. Agreement is poor. The bad results are partly due to the paucity of data points, partly to the irregular shape of the data, and partly to the low number of coefficients in the fitting curve. There were too few points to determine the parameters successfully with a higher order equation using the simple non-linear curve-fitting technique applied here. Better results can be expected even when dealing with fairly complex data, provided that an adequate number of data have been taken over the entire range of interest, and provided a more powerful curve-fitting technique is used, ensuring convergence even with complex data.

The last test was made with data measured by Monod growing Escherichia coli in a medium containing glucose and sorbitil both at concentrations of 100 mg/l.[1] First, the cells used glucose exclusively. Then, after all glucose had been consumed, a short lag and then growth of sorbitol ensued. Monod named this phenomenon of consecutive utilization of substrates with an intervening lag diauxie.[1] Figure 12 and Table VII compare the observed data with values calculated from what will be referred to as the one part model and the two part model. The one part model fitted the data with a single equation. The equation obtained had the form

$$y = \frac{66.374}{1 + \exp(3.2298 - 0.8449t)} \tag{19}$$

The locus of Eq. (19) is the solid curve in Fig.12. As can be seen, the one part model did not give a good fit, showing no plateau at either intermediate or large times. Again, convergence difficulties prevented fitting a higher order equation to this complex data. To increase the quality of the fit, the data were broken into two parts and two separate equations were used to describe the two parts. With this approach, results were very good. For times less than 3.6 hrs, Eq. (20) was applied.

$$y = \frac{29.62}{1+\exp(2.387-.7017t+.6046t^2-1.313t^3+.7225t^4-.1285t^5)}$$

For times greater than 3.6 hrs, Eq. (21) was used to fit the data.

$$y = 29.0+\frac{29.500}{1+\exp(4.0603-.0597(t-3.6)\ -1.626(t-3.6)^2-1.1903(t-3.6)^3)}$$

Excellent agreement was found between model and data, as shown by Table VII and by Fig.12, where the two part model is represented by the dashed curve. The average relative error was 0.78%.

DISCUSSION

Adequate representation of simple batch culture kinetic data may be obtained using the generalized logistic equation. This was borne out by both the results discussed here and a considerable body of other results obtained in analysis of data from a kinetic study of sulfate reduction by bacteria discussed elsewhere.[9] The generalized logistic equation can be used to describe the decrease of substrate concentration with time in a batch culture, although only the description of cell and product concentrations were attempted in this study.

For more complicated data, such as that obtained with the penicillin fermentation, the generalized logistic equation can be used only

when adequate data is available over the entire range of interest. Convergence difficulties may also arise in fitting complex data with the method used here to compute parameters. By using a more sophisticated technique to determine the best values of the fittable parameters, the generalized logistic equation should be able to describe the more complex data, also. With the present approach, good results may be obtained by dividing complex data into two or more regions, as was done with the diauxie data of Monod, but this decreases the convenience of the method somewhat. The representation of the lag phase could be improved, because use of the present method results in a discontinuity at the end of the lag phase. However, this disadvantage is not serious because the rate of product formation at the end of the lag is not generally significant anyway.

Despite the success obtained by choosing a polynomial for F(t), still better results might result from using a different functional form. For example,

$$F(t) = a_o + a_1 t^{a_2} + a_3 e^{a_4 t} \tag{22}$$

will give still greater flexibility for fitting complex batch data. The disadvantage of such a form is the increased difficulty of obtaining good initial estimates of the parameters for use in beginning the iteration. Troubles sometimes arose in using the Chi-square criterion to test goodness of fit for the sulfate reduction system, the primary reason being the difficulty of getting accurate estimates of the standard deviations of the analytical procedures.[9] Accurate standard deviations are also needed if one wishes to weight the data according to statistical criteria. If such information is not available, adequate results can be obtained by using unit weighting factors when data values

do not vary by more than about two orders of magnitude. For wider variations (e.g. total cell count) data points can be weighted in inverse proportion to their magnitude.

CONCLUSIONS

A generalized form of the logistic equation was shown to be useful in the analysis of a variety of batch culture kinetic data. By using a computer program to perform the analysis, the data may be reduced quickly to an equation. The equation may be used to perform manipulations of the data (e.g., differentiation or interpolation) and physical meaning can be derived from the values of some of the parameters in the equation. The equation is a method of storing data concisely. The computer can also be used to prepare graphs comparing the observed data and the fitted curve, but better results could be obtained with complex data if a more powerful numerical method was used to determine the fittable parameters. The generalized logistic equation should prove to be a useful tool in the treatment of batch culture kinetic data.

ACKNOWLEDGEMENTS

The assistance provided by the Math and Computing group at the Lawrence Radiation Laboratory, Berkeley is gratefully acknowledged.

This work was performed under the auspices of the U.S. Atomic Energy Commission.

REFERENCES

1. Monod, J. "Recherches sur la croissance des cultures bacteriennes". Hermann et Cie, Paris (1942).

2. Malek, Ivan and Zdenek Fencl "Theoretical and Methodological of Continuous Culture of Microorganisms", Academic Press, New York (1966).

3. G. Knowles, A.L. Downing, and M.J. Barrett, J. Gen.Microbiol., 38,263 (1965).

4. G. Knowles, Private communication (1967).

5. Verhulst, P. F., Mem. Acad. Roy. Bruxelles, 18, 1 (1844); 20, 1 (1846).

6. Pearl, R. and L. J. Reed, Proc. Nat'l. Acad. Sci., 6, 275 (1920).

7. Dewitt, C. C., Ind. Eng. Chem., 35, 695 (1943).

8. Scarborough, J. B. "Numerical Mathematical Analysis", 4th Ed. p. 478, The Johns Hopkins Press, Baltimore (1958).

9. Edwards, Victor Henry "Analytical Methods in Bacterial Kinetics", Ph.D. Thesis, University of California, Berkeley, California UCRL - 16398, January (1967).

10. Hodgman, Charles D., Robert C. Weast, and Samuel M. Selby "Handbook of Chemistry and Physics", 40th Ed., Chemical Rubber Publishing Co., Cleveland, (1958.)

11. Luedeking, R. L., "The Lactic Acid Fermentation at Controlled pH: Kinetics of the Batch Process and Continuous Flow Theory and Experiment" (Ph.D. Thesis), University of Minnesota, Minneapolis, Minnesota (1956).

12. Hosler, Peter and Marvin J. Johnson, Ind. Eng. Chem. 45, 871 (1953).

Table I. Desirable Properties of a Fitting Equation

1. Capable of representing all phases of batch culture growth and metabolism.

2. Sufficiently flexible to fit many types of data without introducing distortion.

3. At least some of the fittable parameters have direct physical meaning.

4. Continuously differentiable, explicit time derivative.

5. Derivative zero initially and at very large times.

6. Parameters easily evaluated.

7. Easy to use model once parameters determined.

Table II. Data of test with exponential model

Independent variable	Dependent variable		
	Given	Calculated, F(t) first degree	Calculated F(t) fifth degree
0	1.0000	1.0000	1.0000
1	1.6487	1.5838	1.6773
2	2.7183	2.7022	2.6915
3	4.4817	4.5952	4.4136
4	7.3891	7.7706	7.4159
5	12.182	13.019	12.406
6	20.086	21.485	20.194
7	33.115	34.615	32.395
8	54.598	53.794	53.921
9	90.017	79.486	93.892
10	148.41	110.19	138.12
11	148.41	142.24	150.88
12	148.41	171.30	151.47

Table III. Specific growth rates, exponential model

Independent variable	Time derivative of dependent variable		
	Given	Calculated, F(t) first degree	Calculated, F(t) fifth degree
0	0.5	--	--
1	"	.535	.478
2	"	.533	.479
3	"	.529	.510
4	"	.521	.522
5	"	.510	.503
6	"	.491	.474
7	"	.461	.481
8	"	.418	.544
9	"	.360	.526
10	--	.291	.212
11	0	.219	.0162
12	0	.154	.00012

Table IV. Comparison of observed and graphically measured values with computer calculated values (observed data is from run 13 of Luedeking)

Time (hrs)	Observed cell conc. (UOD/ml)	Calculated cell conc. (UOD/ml)	Graphical growth rate (UOD/ml/hr)	Computer calculated growth rate (UOD/ml/hr)
1.0	0.123	0.138	-	0.046
2.0	0.139	0.190	-	0.061
2.5	-	-	0.10	0.077
3.0	0.283	0.268	0.13	0.099
3.5	-	-	0.16	0.130
4.0	0.442	0.400	0.19	0.172
4.5	-	-	0.24	0.228
5.0	-	-	0.30	0.300
5.5	0.82	0.802	0.37	0.389
6.0	-	-	0.45	0.496
6.25	1.12	1.154	-	-
6.50	-	-	0.56	0.618
7.0	-	-	0.71	0.750
7.17	1.77	1.773	-	-
7.50	-	-	0.88	0.885
7.70	2.24	2.23	-	-
8.00	-	-	1.04	1.019
8.25	2.73	2.79	-	-
8.50	-	-	1.17	1.147
9.0	3.77	3.67	1.31	1.271
9.50	-	-	1.43	1.393
9.53	4.43	4.38	-	-
10.00	5.00	5.07	1.55	1.512
10.50	5.72	5.85	1.63	1.612
11.00	6.78	6.67	1.68	1.653
11.50	7.51	7.48	1.56	1.578
12.00	8.23	8.22	1.29	1.340

Table IV (continued)

Time (hrs)	Observed cell conc. (UOD/ml)	Calculated cell conc. (UOD/ml)	Graphical growth rate (UOD/ml/hr)	Computer calculated growth rate (UOD/ml/hr)
12.33	8.63	8.62	-	-
12.50	-	-	1.06	0.965
12.67	8.78	8.95	-	-
13.00	9.27	9.18	0.76	0.564
13.50	9.49	9.38	0.17	0.257
14.00	9.38	9.46	-	-

Table V. Comparison of synthetic data and computer fit values, bell-shaped curve

Independent variable	Dependent variable	
	Original data	Fitted values
0	2.1000×10^6	2.1000×10^6
10.5	"	"
11.0	"	2.5397×10^6
11.5	5.0210×10^6	5.6976×10^6
12.0	1.5620×10^7	1.4531×10^7
12.5	4.2250×10^7	3.6143×10^7
13.0	8.1530×10^7	7.2110×10^7
13.5	1.1090×10^8	1.0516×10^8
14.0	1.2440×10^8	1.2308×10^8
14.5	1.2950×10^8	1.3048×10^8
15.0	1.3140×10^8	1.3317×10^8
15.5	1.3140×10^8	1.3364×10^8
16.0	1.2970×10^8	1.3202×10^8
16.5	1.2260×10^8	1.2492×10^8
17.0	9.0400×10^7	9.6335×10^7
17.5	2.9880×10^7	3.1074×10^7
18.0	2.3310×10^6	2.3806×10^6
18.5	6.8590×10^4	6.5128×10^4

Table VI. Comparison of fitted curve with data of Hosler and Johnson [10]

Time (hours)	Mycelial nitrogen	
	Observed (mg/liter)	Calculated (mg/liter)
0	100	100.0
16	200	235.7
28	-	-
30	900	508.5
40	920	804.3
45	920	871.0
50	-	-
53	1010	1235.1
67	1690	1603.8
75	2000	1738.1
90	1800	1849.4
115	1600	1598.3

Table VII. Comparison of diauxie data of Monod [1] with fitted data

Time (hours)	Data of Monod	Optical density 1-part model	Optical density 2-part model	
0	2.5	2.50	2.500	Part 1
1.00	5.0	5.60	5.100	
1.40	7.8	7.59	7.565	
1.90	11.0	10.92	11.444	
2.10	13.5	12.55	13.179	
2.50	18.0	16.36	17.964	
2.80	23.5	19.68	23.614	
3.15	29.0	24.00	28.820	
3.60	29.5	30.07	29.621	
			29.500	Part 2
4.00	30.0	35.67	30.000	
4.25	32.0	39.11	32.000	
4.60	40.0	43.71	40.000	
4.90	48.0	47.33	48.000	
5.05	53.0	49.00	53.000	
5.25	58.0	51.08	58.000	
5.50	58.5	53.42	58.500	
5.85	58.5	56.23	58.500	
6.10	58.5	57.92	58.500	

APPENDIX I

ESTIMATION OF COEFFICIENTS IN THE GENERALIZED LOGISTIC EQUATION

Accurate initial guesses for the values of the coefficients increase the likelyhood that the iterative curve-fitting technique used will converge to the correct values and reduces the number of iterations required. Assuming that a fairly good estimate can be made for K, we define

$$G(t,y) = \ln\left(\frac{K - y}{y}\right)$$
$$= a_o + a_1 t + a_2 t^2 + a_3 t^3 + a_4 t^4 + \ldots$$

Assume that the initial value $G(0,y_o)$, and three equally spaced values $G(t_1,y_1)$, $G(t_2,y_2)$, and $G(t_3,y_3)$ are known ($t_3 = 3t_1$ and $t_2 = 2t_1$). Then the following equations may be derived to make initial guesses of the coefficients a_o, a_1, etc.

<u>For $F(t) = a_o + a_1 t$</u>

$$a_o = G_o$$
$$a_1 = (G_1 - G_o)/t_1$$

<u>For $F(t) = a_o + a_1 t + a_2 t^2$</u>

$$a_o = G_o$$
$$a_1 = (4(G_1 - G_o) - (G_2 - G_o))/2t_1$$
$$a_2 = ((G_2 - G_o) - 2(G_1 - G_o))/2t_1^2$$

<u>For $F(t) = a_o + a_1 t + a_2 t^2 + a_3 t^3$</u>

Defining $B_1 = G_1 - G_o$

$$B_2 = G_2 - G_o$$
$$B_3 = G_2 - G_o$$

$$a_1 = (3B_1 - 1.5\ B_2 + B_2 + B_3/3)/t_1$$
$$a_2 = (4B_2 - B_3 - 5B_1)/2t_1^{\ 2}$$
$$a_3 = (B_3 - 3B_2 + 3B_1)/6t_1^{\ 3}$$

These equations were used in the computer program to make initial guesses of the coefficients from given equally-spaced data. For fitting of F(t) for fourth and fifth degree cases, zero values of a_4 and a_5 were used for the initial guess. When analyzing a set of data with a lag phase, the foregoing procedure was applied to the data after the time scale had been shifted to place the origin at the end of the lag phase.

APPENDIX II
COMPUTER PROGRAM

The computer program developed to analyze batch culture kinetic data is described by Edwards.[9] A brief description will be given here. Several groups of data may be analyzed in sequence. During analysis of a group, the following procedure is followed: First, the data are read into the computer, along with standard deviations of the data, the duration of the lag phase required, and estimates of dependent variable for use in making estimates of the initial values of the parameters of the fitting equation. The standard deviations are needed to calculate weighting factors for the data in the least squares analysis and for use in the statistical test of goodness of fit of the final equation. Weighting functions are calculated and the data, standard deviations, and the weights of the data are printed out.

If the time lag is not zero, the coordinates are shifted so that in the new coordinate system, $t_1 = t - t_L$, where t_1 = new time scale, t = real time scale, and t_L = length of the lag phase. The value of the dependent variable at $t = t_L$ is assumed to be the average of all values for $t \leq t_L$ and the fitting of the equation to the remaining data is done in the new coordinates.

Estimates of the corrections to the estimated values of the fitting parameters are computed for $F(t_1) = a_o + a_1 t_1$ and the values of the parameters are corrected. This process is continued until the sum of the absolute magnitude of the ratio of the changes in the coefficients to their absolute value is less than .001. The values of the corrections

and the corrected parameters are printed out at each step. At each tenth step, the values of the data and the corresponding values calculated with the fit equation are compared in a table. A maximum of 100 iterations is allowed to reach the desired convergence on the fit parameters. After the fitted values are obtained, the data and the values computed with the fitting equation are compared. Then the calculated derivatives and specific rates are calculated at regular intervals and printed-out also. Finally, a graph is prepared comparing the original data points with the line connecting the values calculated using the fitted equation. The process, starting with estimation of parameters and optimization by iteration, is repeated for third and fifth degree polynomials, provided that the number of data warrant it and that the additional terms are necessary to improve the fit to the data. The next variable in that group of data is then analyzed by the same procedure, until the last variable has been analyzed. The next set of data is then analyzed. Cell concentration data must always be analyzed first because the fitted values of cell concentration are needed in computation of specific rates of change in later variables. The graphs prepared by the computer are very helpful in decisions about the adequacy of the equations fitted with the program. Chi-square values are also computed to assist in judgment of the goodness of fit.

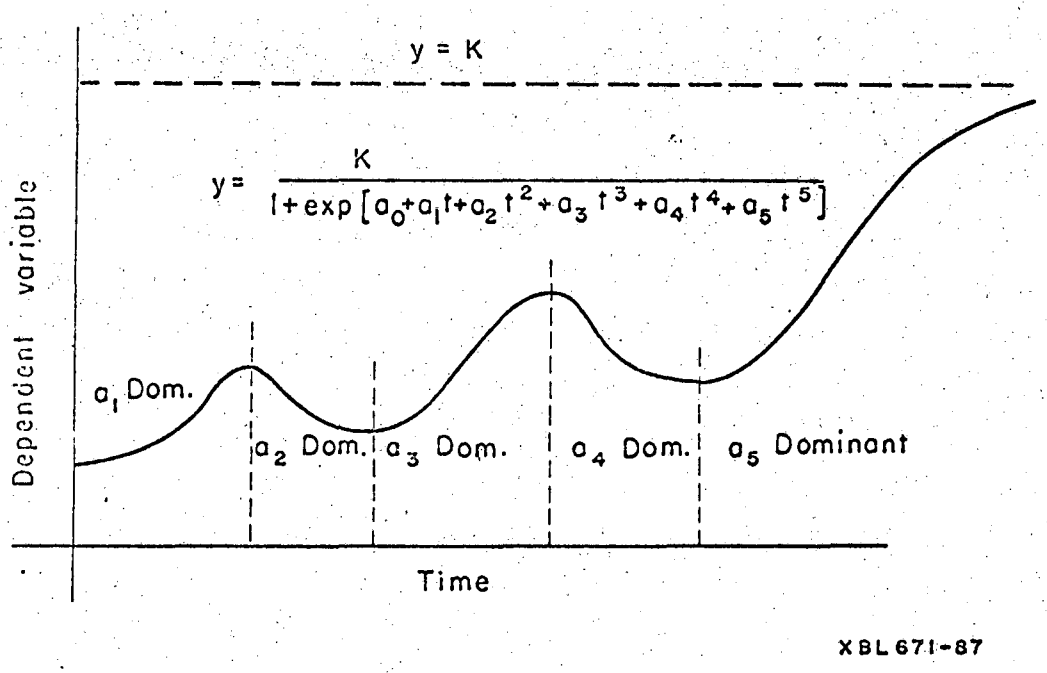
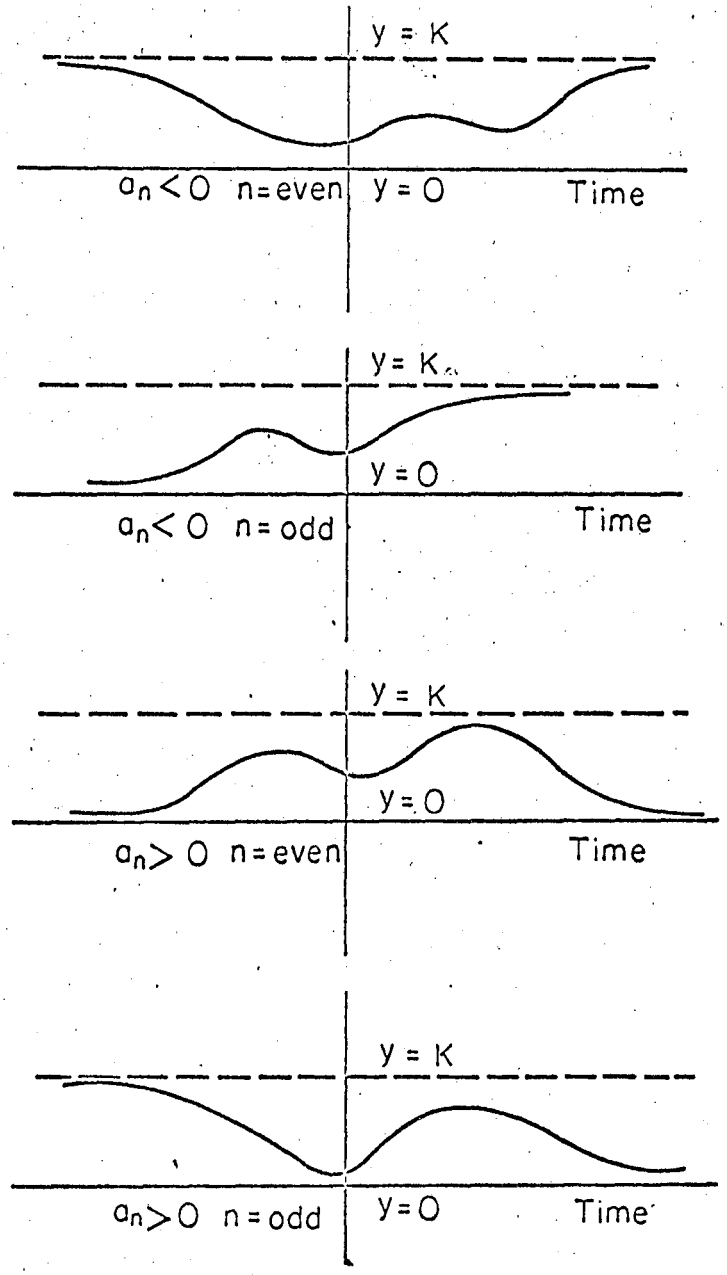


Fig. 1. Typical plot of fifth degree fitting equation versus time.



XBL 671-245

Fig. 2. Possible limiting cases for fitting model at large positive and negative values of independent variable.

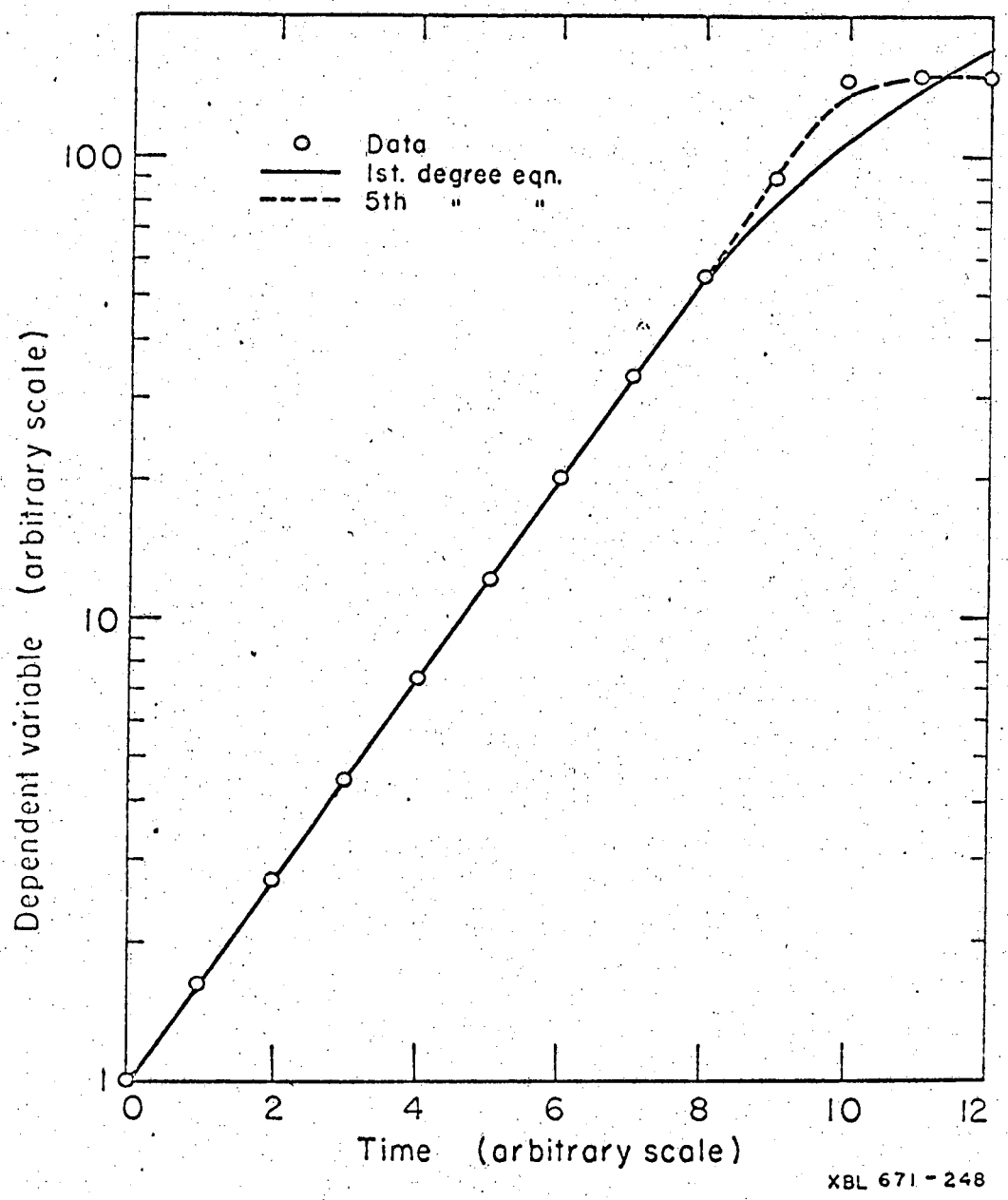


Fig. 3. A comparison of exponential test data with first and fifth degree fitted curves.

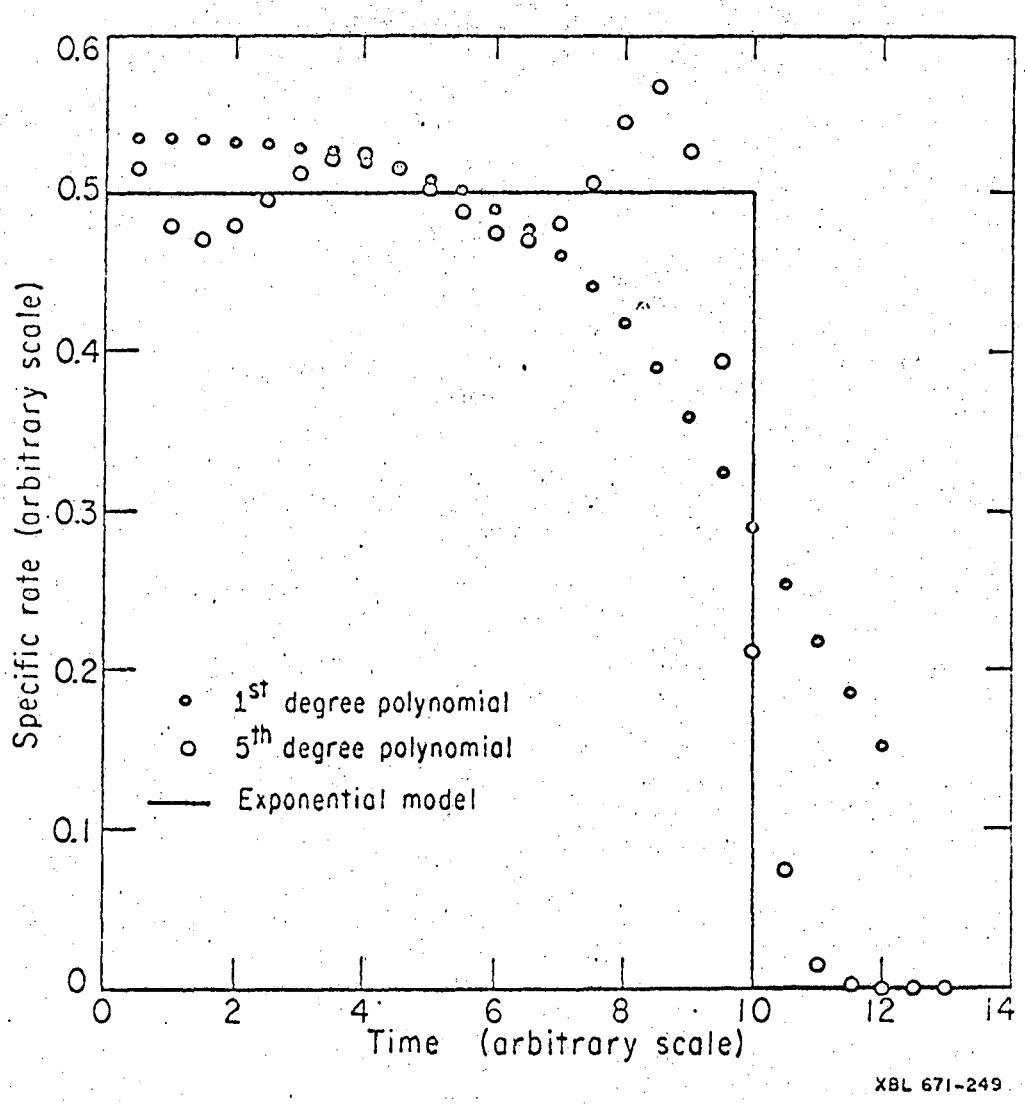


Fig. 4. A comparison of the specific growth rates of the exponential model with values calculated from the first and fifth degree fitted equations.

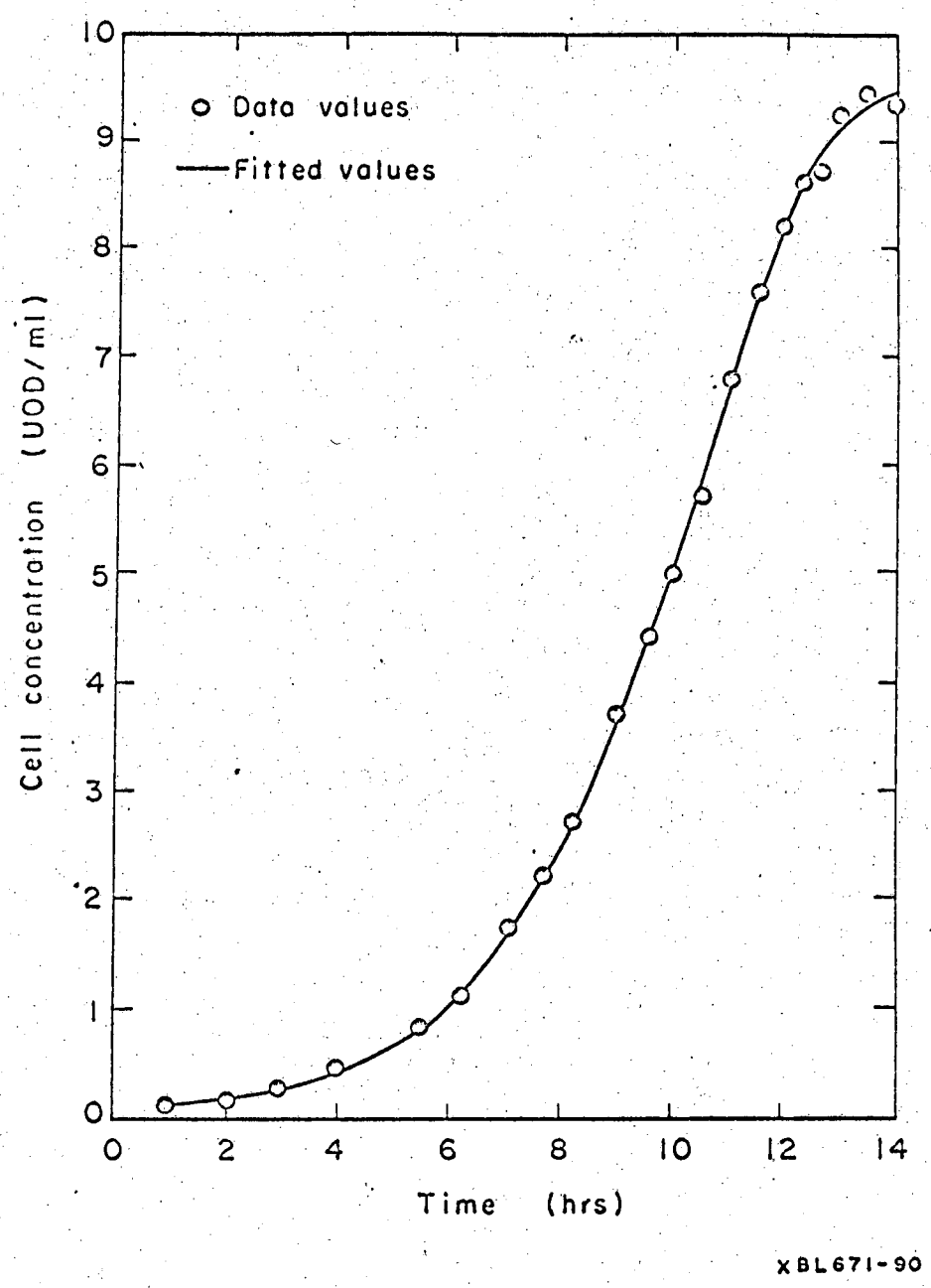


Fig. 5. Comparison of cell concentration data (run 13) of Luedeking[11] with fitted fifth degree curve.

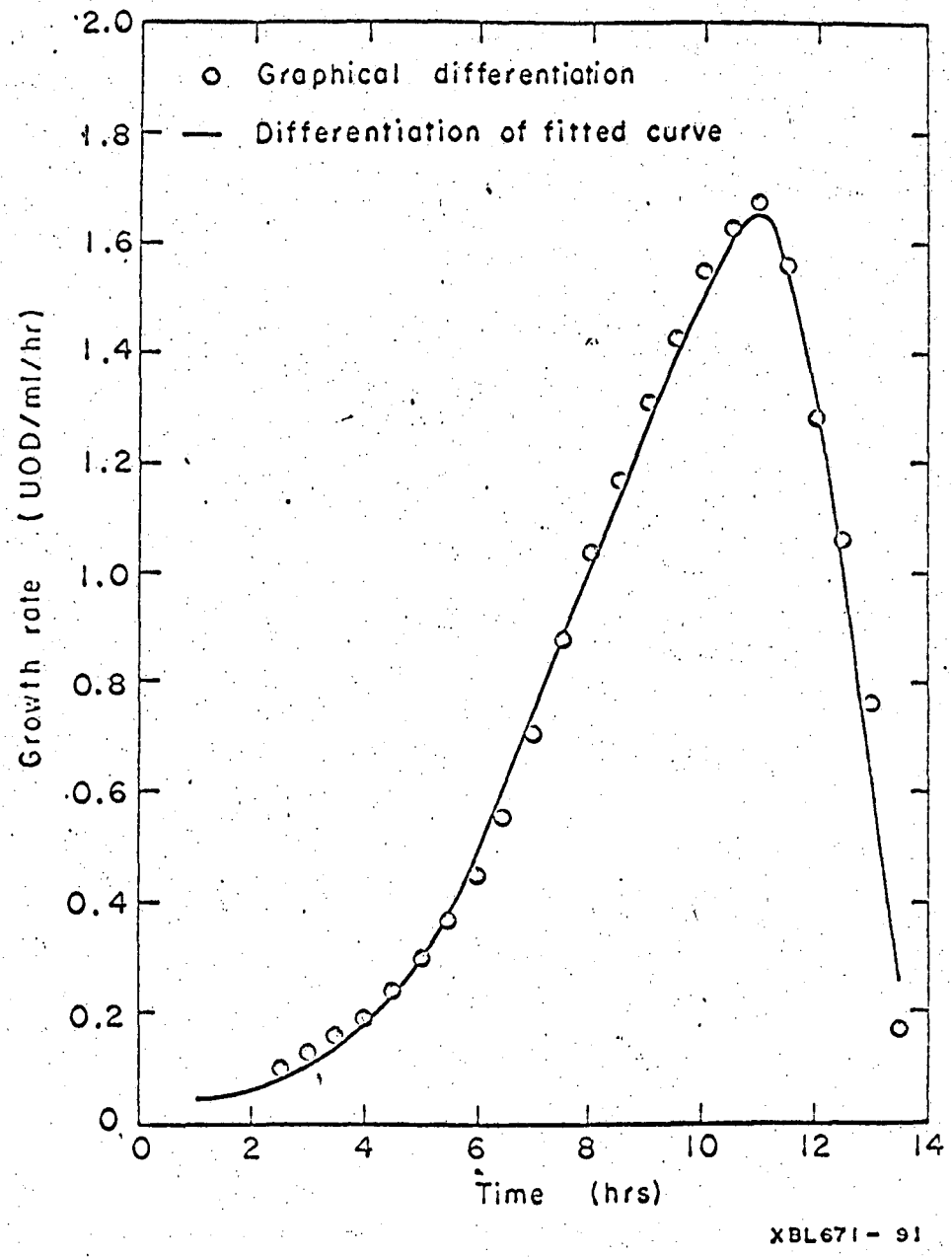


Fig. 6. Comparison of graphically measured growth rate of Luedeking with values calculated as derivatives of the fitted curve.

11

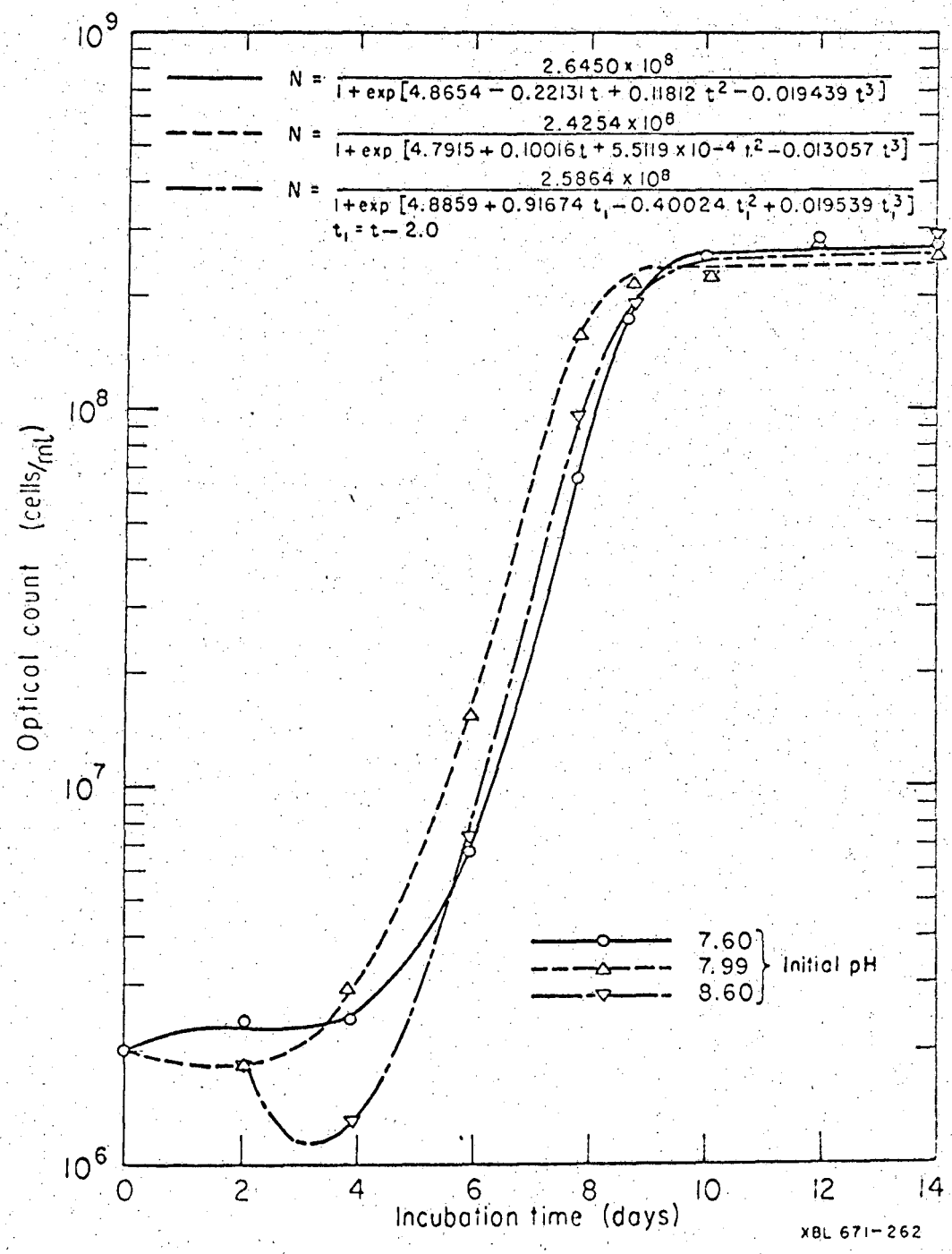


Fig. 7 Effect of the initial pH value on cell growth in batch culture

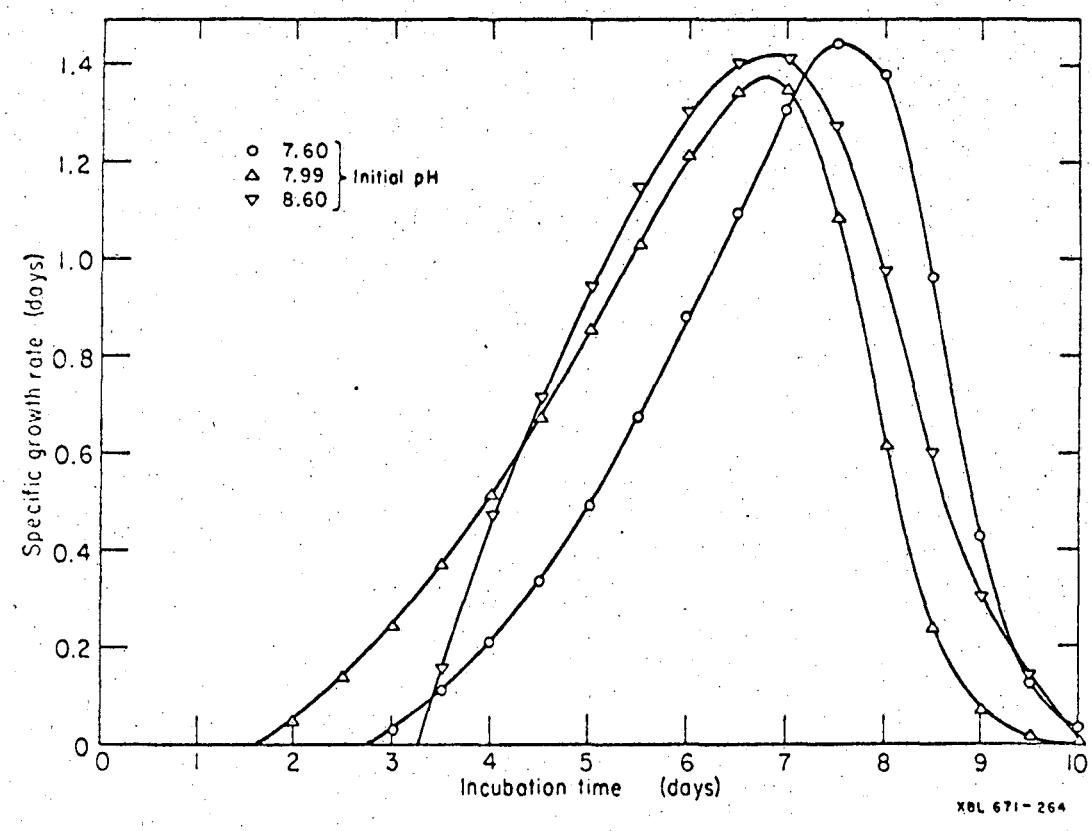


Fig. 8. Variation of specific growth rate with incubation time at different initial pH values

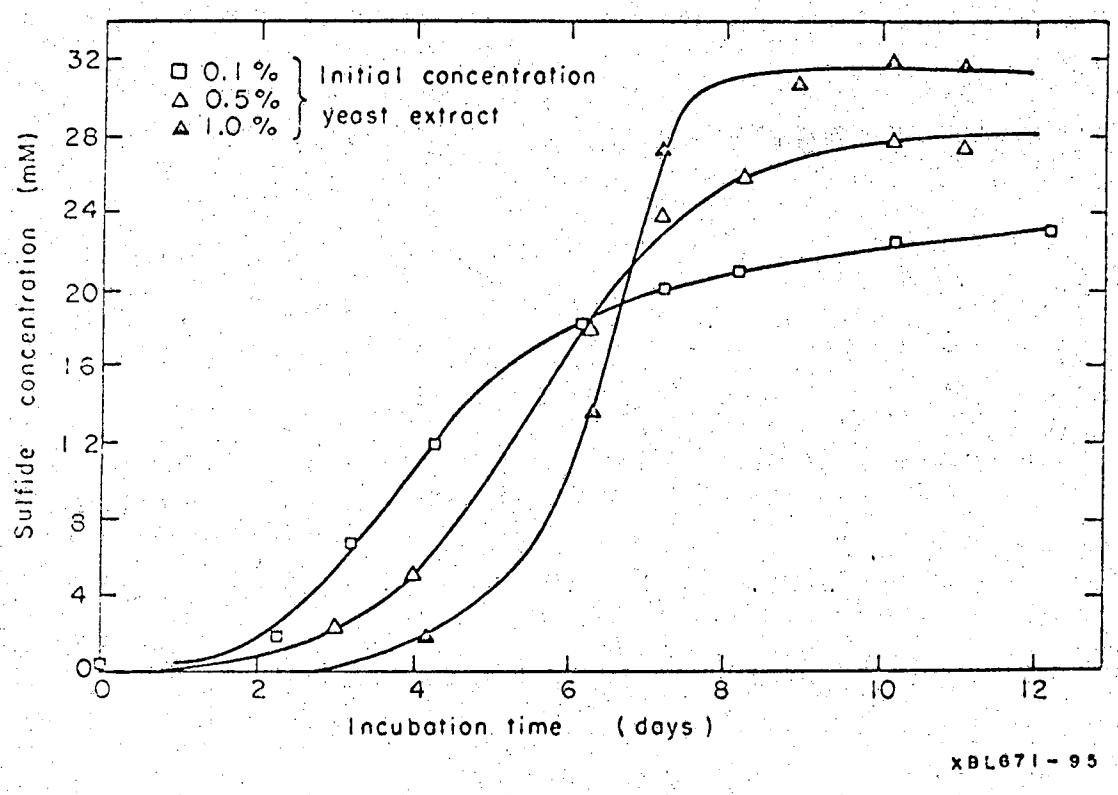


Fig. 9. Effect of the initial concentration of yeast extract on sulfide production in batch culture

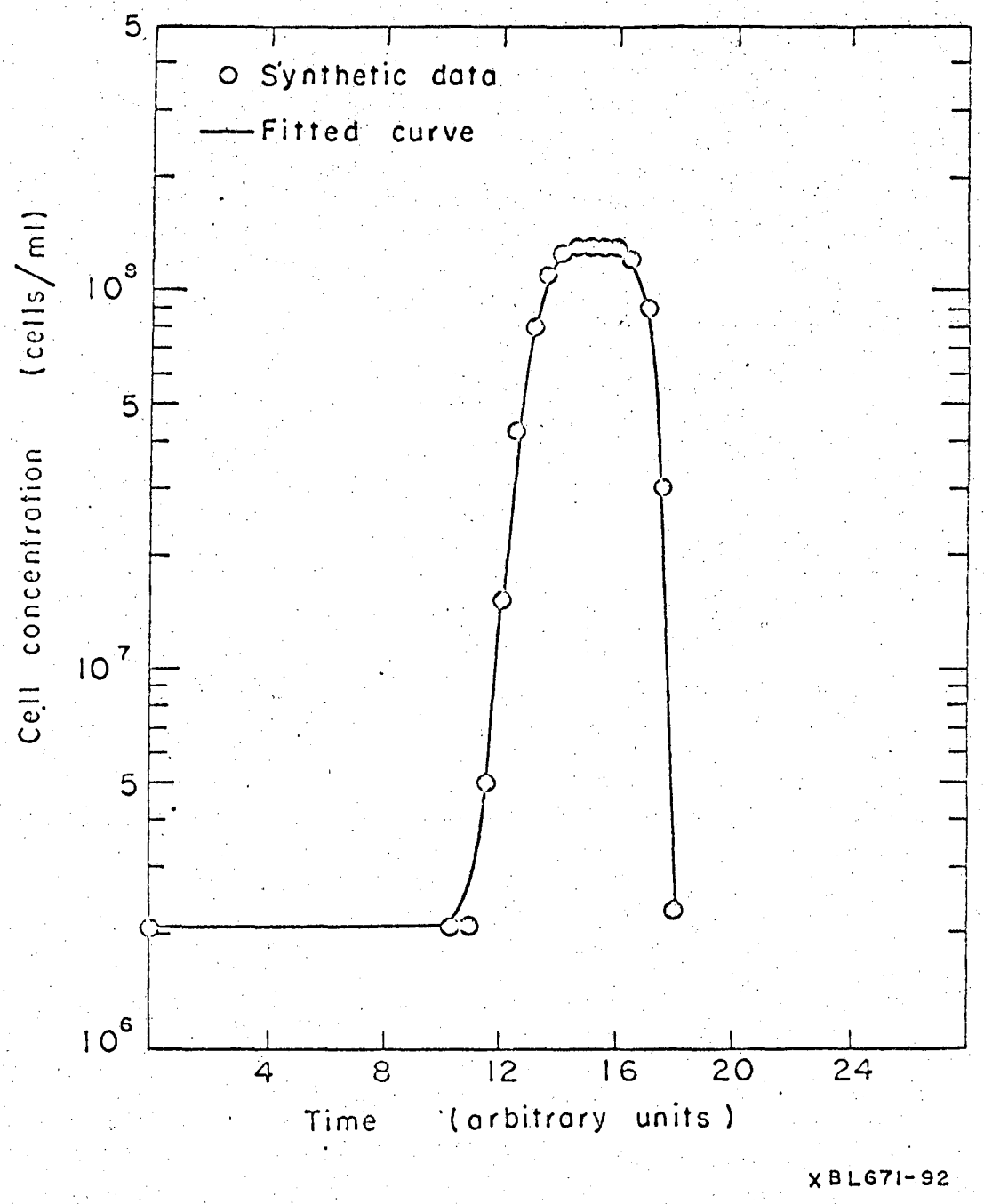


Fig. 10 Comparison of bell-shaped data with fitted equation.

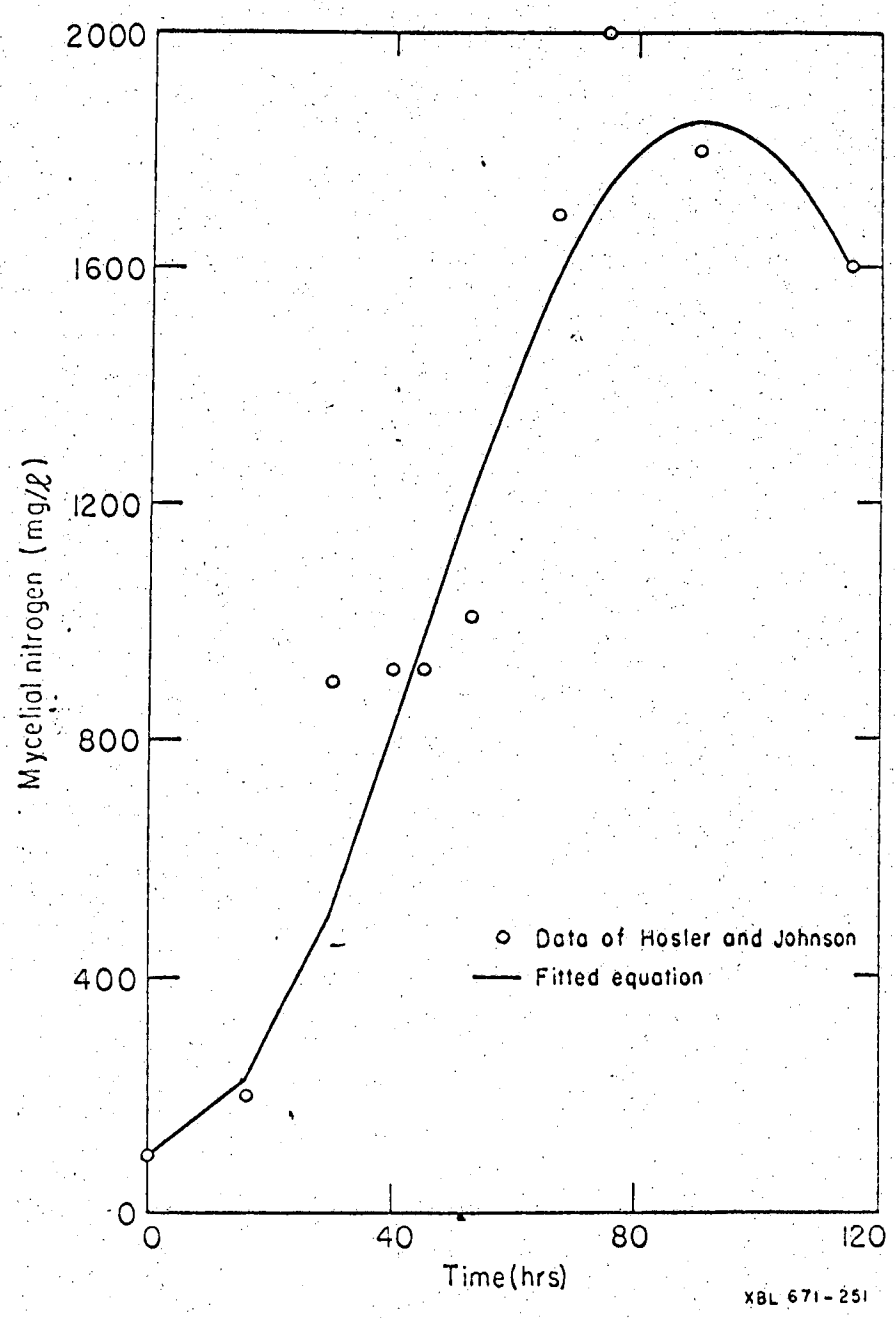


Fig. 11 Mycelial nitrogen concentration in the penecillin fermentation compared with values calculated from the fitted curve.

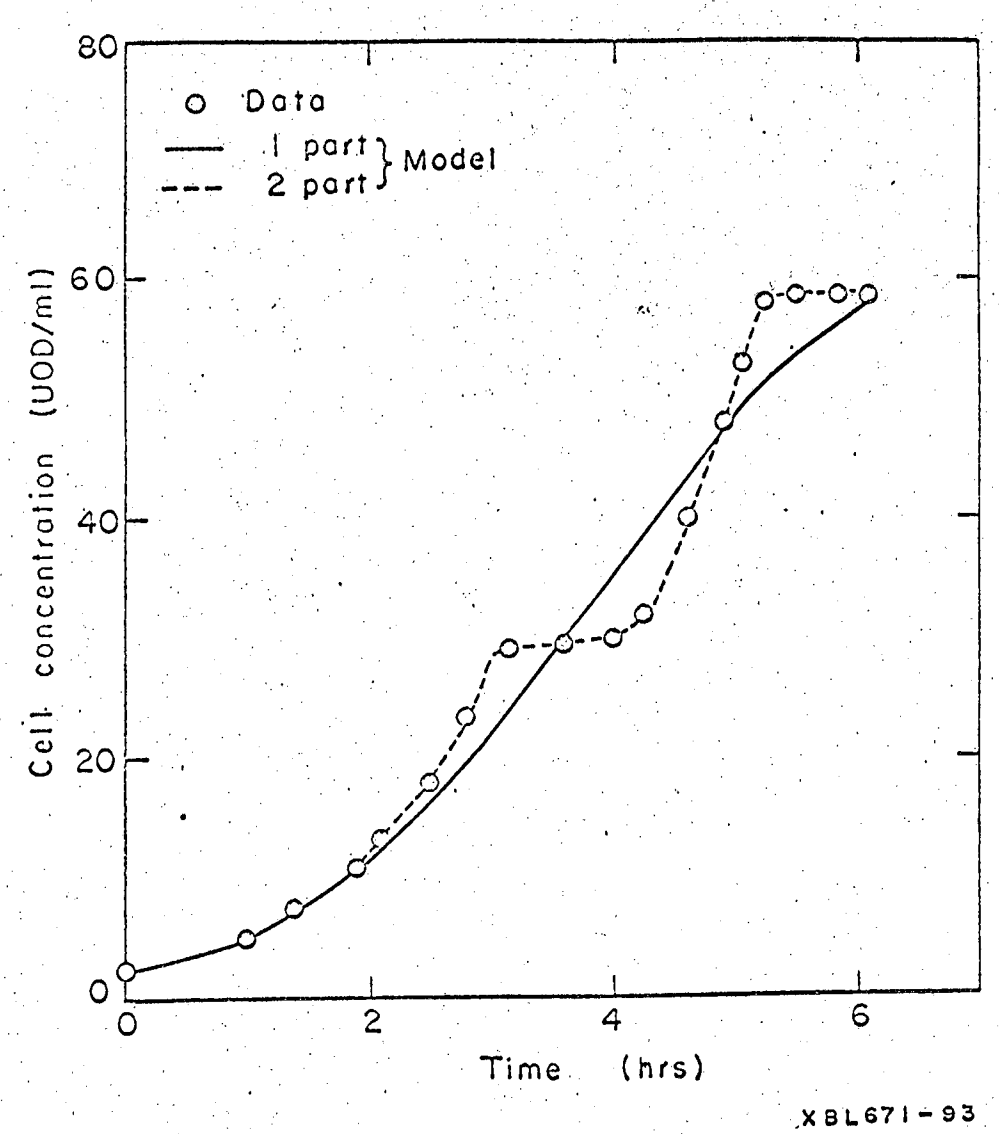


Fig. 12 Cell concentration versus time in diauxic batch culture compared with one-part and two-part fitted curves.

This report was prepared as an account of Government sponsored work. Neither the United States, nor the Commission, nor any person acting on behalf of the Commission:

A. Makes any warranty or representation, expressed or implied, with respect to the accuracy, completeness, or usefulness of the information contained in this report, or that the use of any information, apparatus, method, or process disclosed in this report may not infringe privately owned rights; or

B. Assumes any liabilities with respect to the use of, or for damages resulting from the use of any information, apparatus, method, or process disclosed in this report.

As used in the above, "person acting on behalf of the Commission" includes any employee or contractor of the Commission, or employee of such contractor, to the extent that such employee or contractor of the Commission, or employee of such contractor prepares, disseminates, or provides access to, any information pursuant to his employment or contract with the Commission, or his employment with such contractor.