

# Statistical Analysis of Nonlinear Parameter Estimation for Monod Biodegradation Kinetics Using Bivariate Data

Christopher D. Knightes, Catherine A. Peters

Department of Civil and Environmental Engineering, Princeton University, Princeton, New Jersey 08544, USA; telephone: (609)-258-5645; fax: (609)-258-2799; e-mail: cap@princeton.edu

Received 5 June 1999; accepted 7 November 1999

**Abstract:** A nonlinear regression technique for estimating the Monod parameters describing biodegradation kinetics is presented and analyzed. Two model data sets were taken from a study of aerobic biodegradation of the polycyclic aromatic hydrocarbons (PAHs), naphthalene and 2-methylnaphthalene, as the growth-limiting substrates, where substrate and biomass concentrations were measured with time. For each PAH, the parameters estimated were:  $q_{\max}$ , the maximum substrate utilization rate per unit biomass;  $K_S$ , the half-saturation coefficient; and  $Y$ , the stoichiometric yield coefficient. Estimating parameters when measurements have been made for two variables with different error structures requires a technique more rigorous than least squares regression. An optimization function is derived from the maximum likelihood equation assuming an unknown, nondiagonal covariance matrix for the measured variables. Because the derivation is based on an assumption of normally distributed errors in the observations, the error structures of the regression variables were examined. Through residual analysis, the errors in the substrate concentration data were found to be distributed log-normally, demonstrating a need for log transformation of this variable. The covariance between  $\ln C$  and  $X$  was found to be small but significantly nonzero at the 67% confidence level for NPH and at the 94% confidence level for 2MN. The nonlinear parameter estimation yielded unique values for  $q_{\max}$ ,  $K_S$ , and  $Y$  for naphthalene. Thus, despite the low concentrations of this sparingly soluble compound, the data contained sufficient information for parameter estimation. For 2-methylnaphthalene, the values of  $q_{\max}$  and  $K_S$  could not be estimated uniquely; however,  $q_{\max}/K_S$  was estimated. To assess the value of including the relatively imprecise biomass concentration data, the results from the bivariate method were compared with a univariate method using only the substrate concentration data. The results demonstrated that the bivariate data yielded a better confidence in the estimates and provided additional information about the model fit and model adequacy. The combination of the value of the bivariate data set and their nonzero covariance justifies the need for maximum likelihood esti-

mation over the simpler nonlinear least squares regression. © 2000 John Wiley & Sons, Inc. *Biotechnol Bioeng* 69: 160–170, 2000.

**Keywords:** nonlinear parameter estimation; maximum likelihood; Monod kinetics; biodegradation kinetics; polycyclic aromatic hydrocarbons; naphthalene; 2-methylnaphthalene

## INTRODUCTION

The Monod equation is commonly used to model the kinetics of biodegradation of organic compounds and associated biomass growth. The depletion rate of the substrate [Eq. (1)] is described by the Monod parameters,  $q_{\max}$ , the maximum substrate utilization rate per unit biomass (mg substrate/mg biomass [protein]/hour), and  $K_S$ , the half-saturation coefficient (mg/L). Biomass growth rate [Eq. (2)] is described by the same parameters, and by the yield coefficient,  $Y$  (mg protein/mg substrate), and the endogenous decay rate,  $b$  ( $\text{h}^{-1}$ ). The differential equations for time dependence of concentration of a growth-limiting substrate and biomass growth are:

$$\frac{dC}{dt} = -q_{\max} \frac{C}{K_S + C} X \quad (1)$$

$$\frac{dX}{dt} = q_{\max} Y \frac{C}{K_S + C} X - bX \quad (2)$$

where the dependent variables are substrate concentration,  $C$  (mg/L), and biomass concentration,  $X$  (mg protein/L), and the independent variable is time,  $t$  (h). Parameter estimation is challenging for these equations because they constitute a set of coupled nonlinear differential equations that exhibits an asymptotic behavior.

The asymptotic nature of the Monod formulation results in an interesting complication for parameter estimation. At very high substrate concentrations,  $C \gg K_S$ , the microorganisms are utilizing the substrate at the fastest rate possible, denoted by  $q_{\max}$ , and the model simplifies to a single-parameter, zero-order growth model. As the substrate concentration decreases, the model transitions into a mixed-

Correspondence to: C. Peters

Contract grant sponsors: National Science Foundation; Department of Education (graduate assistantships in Areas of National Need Program)

Contract grant numbers: EAR-9708406; P200A80804

order rate. As the substrate reaches a low concentration,  $C \ll K_S$ , and biomass growth is slow, the model reduces to a single-parameter model with first-order rate constant proportional to  $q_{\max}/K_S$ . For sparingly soluble compounds, such as PAHs, it can be difficult to estimate uniquely the parameters  $q_{\max}$  and  $K_S$ . Ellis et al. (1996) performed a sensitivity analysis for estimating these parameters and suggested that when the ratio of initial substrate,  $C_0$ , to  $K_S$  is less than 0.1, then  $q_{\max}$  and  $K_S$  cannot be estimated uniquely. They suggested that when  $C_0:K_S$  is 1.0 or higher, there is good separation of these two parameters. These ratios are useful guidelines, but the ability to predict unique estimates of  $q_{\max}$  and  $K_S$  is dependent on the number of observations and the precision of the measurements (Robinson, 1998).

Different methods have been presented in the literature for estimation of these parameters. A common method is to transform the nonlinear Monod equations into a linearized form, and then to estimate parameters using linear regression techniques. Although this approach is simple, the assumptions of normally distributed errors inherent in linear regression may not be valid because the linearization alters the error distributions (Leatherbarrow, 1990). The use of this simplification has been largely due to the computational difficulty involved in nonlinear parameter estimation. Nonlinear regression involves an iterative method for optimal values, unlike the closed-form analytical equations that result when the model is linear in the parameters. With the recent advances in computing capabilities, the use of nonlinear parameter estimation techniques has become more feasible (Leatherbarrow, 1990).

Another method of simplification involves an integrated form of the Monod equation. By assuming that the endogenous biomass decay is negligible, the change in biomass concentration is related to substrate depletion solely through stoichiometry. An algebraic relation for the biomass growth is substituted into the substrate equation, which is then integrated. This reduces the formulation to a univariate algebraic equation in which a nonlinear least squares regression is employed (Ong, 1983; Robinson and Tiedje, 1983; Smith et al., 1998). To use the integrated Monod equation, the initial biomass concentration is either measured or estimated as an additional parameter (Guha and Jaffe, 1996; Smith et al., 1997). With sparingly soluble compounds, such as PAHs, the endogenous decay rate can be significant relative to the biodegradation rate, making this method inappropriate.

General least squares regression (for linear or nonlinear models) is derived from the maximum likelihood method under the assumption that the variance is constant and the errors are normal for all observations. In the case in which two types of data are simultaneously fit to a model with common parameters, general least squares regression should not be used if the errors in the two data types cannot be assumed to be identical. Weighted least squares regression may be applicable in this case but this does not allow for the possibility of covariance between the errors in the two data types. When two dependent variables have been

measured and there is the possibility of covariance between them, the maximum likelihood method is the appropriate nonlinear parameter estimation technique. Robinson (1985) stated that maximum likelihood estimation is superior to least squares regression, but is difficult to implement because the covariance matrix of the measurement errors must be known a priori. However, the covariance matrix does not need to be known. Rather, it can be estimated along with the parameters (Bard, 1974).

Many approaches and assumptions have been employed by investigators who have fit Monod parameters to bivariate data. Sleep and Mulcahy (1998) used Bayesian parameter estimation, which incorporates a prior distribution function along with the likelihood equation, for estimating the biokinetic parameters for an organic substrate in an unsaturated soil. They measured carbon dioxide production and substrate (toluene) depletion as a function of time, but they measured biomass concentration only at the onset of the experiment. Sommer et al. (1995) used the maximum likelihood method and measured both the biomass and substrate (toluene) concentration with time. The responses of both variables were assumed to be normally distributed with no covariance. The variance was assumed to be the same for both types of data and was estimated along with the parameters. The variance was divided by a weighting factor that was assumed to be proportional to the mean of each variable type. This effectively weights the residuals such that they are assumed to be log-normally distributed. The variance associated with the biomass was additionally weighted by a factor of three, which the investigators believed to be a reasonable value because the biomass measurements were less precise than the substrate measurements. Guha and Jaffe (1996) measured substrate (phenanthrene) depletion and carbon dioxide production. They used the maximum likelihood method to derive an objective function assuming normally distributed errors in the observations and an unknown covariance matrix.

The development of different approaches over the years to estimate Monod parameters demonstrates its inherent difficulty. This article addresses the problem of nonlinear parameter estimation for bivariate biodegradation rate data by presenting a statistical analysis of a maximum likelihood estimation approach. Two model data sets were obtained from experiments of aerobic biodegradation using naphthalene (NPH) and 2-methylnaphthalene (2MN) as growth-limiting substrates. For each experiment, measurements were made of both substrate and biomass concentrations. To validate the necessary assumptions regarding normality, independence, and constancy of errors within each data set, we examined the need for data transformation and performed a rigorous analysis of the regression residuals. The use of the maximum likelihood approach, as opposed to a least squares regression approach, eliminates the need to assume constancy of the variances and absence of covariance across the two data types. For each model data set, optimal Monod parameters were estimated along with estimates of the covariance matrix of the regression variables.

The issue of uniqueness of the  $q_{\max}$  and  $K_S$  estimates was addressed through a three-dimensional (3D) response surface technique relating the parameter estimation objective function to values of the parameters. Because of the nonlinearities in the response surfaces, it was necessary to transform some of the parameters to approximate their confidence limits and joint confidence regions. Finally, the results from the bivariate regression were compared with the results from a univariate regression using only substrate concentration data. This comparison examined the extent to which the relatively less precise biomass concentration data provide valuable information about the model parameters.

## MATERIALS AND METHODS

### Experimental Data

To determine the endogenous decay rate constant, an independent batch experiment was performed. The experiment consisted of a continually mixed batch reactor containing biomass in a substrate-free buffer solution. Triplicate samples of biomass concentration were measured at four different timepoints over 140 h.

The experiments for measurement of biodegradation rates for the two model substrates are presented here in brief. Details of the experiments have been presented elsewhere (Knights, 2000). For each PAH, a biotic series and an abiotic series of experiments were run in parallel. Each series consisted of 15 independent reactor vessels containing nutrient buffer stock solution with a PAH concentration near aqueous solubility. The biotic series was inoculated with a PAH-degrading consortium. Each reactor was sealed, but contained sufficient oxygen to maintain aerobic conditions throughout the experiment. Each system was well mixed throughout the experiment via magnetic stir bars.

Fifteen minutes after the initiation of the experiments, the first vessel in each series was sacrificed for measurement of sorption to the apparatus and to biorelated material. Fifteen minutes was assumed to allow enough time to account for instantaneous sorption without significant abiotic losses or biodegradation. From the remaining vessels, samples were taken at time intervals dictated by the apparent rate of biodegradation. The PAH concentration was measured using HPLC followed by UV/FLD detection. For the biotic experiments, the biomass concentration was measured as equivalent concentration of bovine gamma globulin protein using the Bradford method. All observations were made by taking the average of triplicate samples to reduce the measurement error. The procedure resulted in a total of 15 independent observations of  $C$  and  $X$  for each substrate. The initial conditions for the experiments were  $C_{0,NPH} = 19.4$  mg/L,  $C_{0,2MN} = 9.85$  mg/L, and  $X_0 = 0.053$  mg protein/L.

### Model Formulation

The system is modeled by combining the Monod equations [Eqs. (1) and (2)] with endogenous decay, time-dependent

abiotic losses, and instantaneous partitioning. The endogenous rate of decay, characterized by rate coefficient  $b$  ( $\text{h}^{-1}$ ), is subtracted from the growth rate of the biomass to account for the energy lost for cell maintenance. Abiotic losses that occur over time are modeled using a first-order rate,  $k_a$  ( $\text{h}^{-1}$ ). Because NPH and 2MN are fairly hydrophobic compounds, the amount of partitioning out of the aqueous phase is relatively important. Therefore, parameters  $K_a$  (dimensionless) and  $K_b$  (L/mg protein) are introduced to account for the sorption to the apparatus and to any biorelated material, respectively.  $K_a$  and  $K_b$  are used to model the instantaneous sorption, whereas  $k_a$  is used to model any abiotic losses that occur throughout the duration of the experiment. Parameters  $K_a$ ,  $K_b$ ,  $k_a$ , and  $b$  were estimated via independent experiments. The mass balance equations are:

$$(1 + K_a + K_b X) \frac{dC}{dt} = -q_{\max} \frac{C}{K_S + C} X - k_a C - K_b C \frac{dX}{dt} \quad (3)$$

$$\frac{dX}{dt} = q_{\max} Y \frac{C}{K_S + C} X - bX \quad (4)$$

The data set is comprised of the measured variables,  $C$  and  $X$ , measured over the independent variable,  $t$ . The parameter vector,  $\underline{\theta} = [\ln q_{\max}, \ln K_S, Y]$ , was estimated using an objective function derived from the maximum likelihood equation.

### Maximum Likelihood Estimation

The likelihood function,  $L(\underline{\theta}, \underline{Y})$ , is defined as the product of the probabilities of all observations for a given parameter vector,  $\underline{\theta}$ , and a given covariance matrix,  $\underline{Y}$ . The maximum likelihood estimate for  $\underline{\theta}$  and  $\underline{Y}$  is determined by finding the maximum value for  $L(\underline{\theta}, \underline{Y})$  (Bevington and Robinson, 1992). To derive  $L(\underline{\theta}, \underline{Y})$ , the mathematical characteristics of the density functions describing the errors in the observations must be known. Assuming the errors are independent, normally distributed, and within each variable type the variance is constant, then the log of the likelihood equation is written as (Bard, 1974):

$$\log L(\underline{\theta}, \underline{Y}) = \frac{nm}{2} \log 2\pi - \frac{n}{2} \log |\underline{Y}| - \frac{1}{2} \text{Tr}[\underline{Y}^{-1} \underline{M}(\underline{\theta})] \quad (5)$$

where  $n$  is the number of observations,  $m$  is the number of independent variables,  $\text{Tr}$  denotes the trace of the matrix,  $|\underline{Y}|$  denotes the determinant of matrix  $\underline{Y}$ , and  $\underline{M}(\underline{\theta})$  is the moment matrix of the residuals:

$$\underline{M}(\underline{\theta}) \equiv \sum_{i=1}^n \underline{e}_i(\underline{\theta}) \underline{e}_i^T(\underline{\theta}) \quad (6)$$

where the superscript,  $T$ , denotes the transpose of the matrix. If there are two measured variables,  $Y_1$  and  $Y_2$ , then the vector of the residuals for observation  $i$  is defined as:

$$\underline{e}_i = [e_{Y_{1,i}}, e_{Y_{2,i}}] = [Y_{1,i} - \hat{Y}_{1,i}, Y_{2,i} - \hat{Y}_{2,i}] \quad (7)$$

Where  $\hat{\cdot}$  denotes the predicted value associated with an observation. The covariance matrix is:

$$\underline{V} = \begin{bmatrix} \sigma_{Y_1}^2 & \sigma_{Y_1 Y_2}^2 \\ \sigma_{Y_1 Y_2}^2 & \sigma_{Y_2}^2 \end{bmatrix} \quad (8)$$

Maximizing  $\log L(\underline{\theta}, \underline{Y})$  is equivalent to minimizing:

$$\Phi(\underline{\theta}) = \frac{n}{2} \log|\underline{V}| + \frac{1}{2} Tr[\underline{V}^{-1} \underline{M}(\underline{\theta})] \quad (9)$$

which becomes the parameter estimation objective function.

If only one variable is measured and the variance is constant, then Eq. (9) reduces to unweighted least squares regression:

$$\Phi(\underline{\theta}) = \sum_{i=1}^n (Y_{1,i} - \hat{Y}_{1,i})^2 \quad (10)$$

When two variables are measured, and constant variance with no covariance is assumed for all observations, then Eq. (9) reduces to least squares regression with weighting by the variance of the measured variable:

$$\Phi(\underline{\theta}) = \frac{1}{\sigma_{Y_1}^2} \sum_{i=1}^n (Y_{1,i} - \hat{Y}_{1,i})^2 + \frac{1}{\sigma_{Y_2}^2} \sum_{i=1}^n (Y_{2,i} - \hat{Y}_{2,i})^2 \quad (11)$$

where  $\sigma_Y^2$  is the variance of the given measured variable. When the variance cannot be assumed constant, then Eq. (9) becomes the least squares regression with weighting by the variance of each variable's observation:

$$\Phi(\underline{\theta}) = \sum_{i=1}^n \frac{(Y_{1,i} - \hat{Y}_{1,i})^2}{\sigma_{Y_{1,i}}^2} + \sum_{i=1}^n \frac{(Y_{2,i} - \hat{Y}_{2,i})^2}{\sigma_{Y_{2,i}}^2} \quad (12)$$

In the event that the variance can be assumed constant within each variable type, but the covariance of the variables cannot be assumed to be negligible, then the objective function becomes:

$$\Phi(\underline{\theta}) = \frac{n}{2} \log|\underline{M}(\underline{\theta})| \quad (13)$$

and the covariance is estimated from the moment matrix by:

$$\hat{\underline{V}}(\underline{\theta}) = \frac{1}{n - l/m} \underline{M}(\underline{\theta}) \quad (14)$$

where  $l$  is the number of estimated parameters (Bard, 1974). Because there is no reason to assume that covariance is negligible between biomass and substrate concentrations, Eq. (13) defines the objective function for the estimation of the Monod parameters.

## PARAMETER ESTIMATION AND STATISTICAL ANALYSIS

### Abiotic Parameter Estimation

The abiotic loss rate constant was independently determined in the abiotic systems, using:

$$\frac{dC_a}{dt} = -k_a C_a \quad (15)$$

where  $C_a$  (mg/L) is the substrate concentration in the abiotic systems. The  $\ln C_a$  vs.  $t$  data were fit to the integrated form of this equation using unweighted least squares linear regression. The regression estimated  $k_a = 0.0148 \text{ h}^{-1}$  for NPH and  $k_a = 0.0095 \text{ h}^{-1}$  for 2MN. For both NPH and 2MN,  $k_a$  was found to be significantly different from zero at the 95% confidence level. Therefore, the abiotic loss must be accounted for in the modeling of these systems.

### Endogenous Decay Coefficient

The endogenous decay coefficient was independently determined using first-order rate kinetics as:

$$\frac{dX}{dt} = -bX \quad (16)$$

Eq. (16) was integrated and the biomass data were transformed into log space. An unweighted least squares linear regression was used to estimate  $b$  with a 95% confidence level as  $b = 0.0042 \pm 0.0019 \text{ h}^{-1}$ .

### Partition Coefficients

The extent of partitioning to the apparatus is estimated from the abiotic experimental system, and the extent of additional partitioning to biorelated material is estimated from the biotic experimental system. Partitioning was assumed to be instantaneous and to follow a linear isotherm. The abiotic partitioning coefficient,  $K_a$ , is estimated as:

$$K_a = \frac{C_T - C_{15,a}}{C_{15,a}} \quad (17)$$

where  $C_T$  is the initial mass of substrate present in the system divided by the system volume, and  $C_{15,a}$  is the aqueous-phase concentration in the abiotic system at  $t = 15$  min. Assuming that the amount of biological activity is negligible in the first 15 minutes, the additional partitioning due to sorption to biorelated material is estimated as:

$$K_b = \frac{1}{X} \left( \frac{C_T - K_a C_{15,b} - C_{15,b}}{C_{15,b}} \right) \quad (18)$$

Where  $C_{15,b}$  is the concentration of the biotic system at  $t = 15$  min. The sorption parameters were estimated to be  $K_a = 0.077$  and  $K_b = 1.36 \text{ L/mg}$  for NPH and  $K_a = 0.199$  and  $K_b = 0.291 \text{ L/mg}$  for 2MN.

### Parameter Estimation

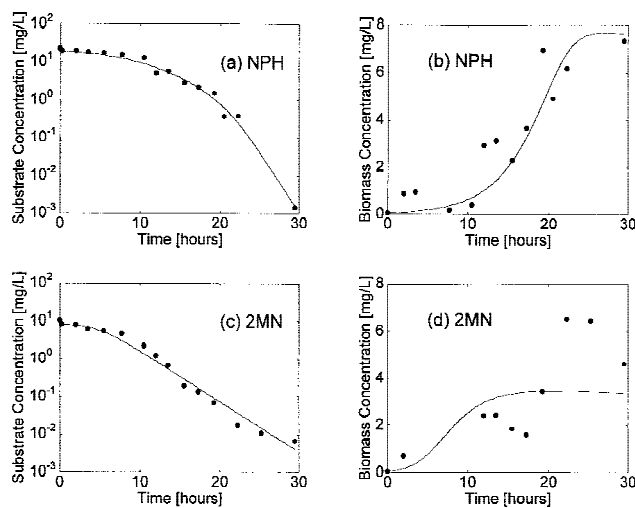
The parameterization approach involves a search algorithm to find where the objective function takes its minimum value. To determine  $\Phi$  [Eq. (13)] a subroutine was called that solves the model equations for a given value of  $\underline{\theta}$ . The

equations were solved numerically using a fourth-order Runge–Kutta algorithm with a time step of 5 min.

The asymptotic nature of the Monod model produces a highly nonlinear relationship between the objective function and the parameters. The irregularities in this relationship hinder the success of standard search algorithms. We found it necessary to use a combination of methods to minimize the objective function. A starting point was selected by determining the minimum  $\Phi$  value on a cubic lattice of eight points in the  $q_{\max}$ – $K_S$ – $Y$  space. The points used to form the grid consisted of  $q_{\max} = 0.6$  and  $6$  mg/mg h<sup>-1</sup>,  $K_S = 1$  and  $10$  mg/L, and  $Y = 0.1$  and  $0.5$  mg/mg, selected to cover an order of magnitude for each parameter encompassing literature values. We used a standard search algorithm to minimize simultaneously the objective function over the three parameters. The optimization algorithm *fmins* (MATLAB, version 5.2.0, The Mathworks, Inc.), which uses the Nelder–Mead type simplex direct search method, was used and required a run time of roughly 20 min. It was found that a transformation of  $q_{\max}$  and  $K_S$  into logarithmic space produced a less skewed objective function surface and enhanced the efficiency and success of the search algorithm. The next step was to verify that the resulting estimate of  $\theta$  was a true minimum of the objective function. Using the estimated value for  $Y$ , a response surface of the objective function in the  $\ln q_{\max}$ – $\ln K_S$  space was plotted, including  $K_S$  values up to an order of magnitude larger than the largest substrate concentration measurement. For NPH, the estimated minimum was indeed found to be unique. For 2MN, the estimated minimum was erroneous. The absence of a minimum in the objective function surface indicated that unique estimates of  $q_{\max}$  and  $K_S$  could not be obtained with the available experimental data. In this case, the data were fit to a reformulated model in which the biodegradation term is first order. The optimization algorithm was used to find values of  $q_{\max}/K_S$  and  $Y$  that minimize the objective function.

For NPH, the minimum for the optimization function was  $\hat{q}_{\max} = 0.636$  mg/mg h<sup>-1</sup>,  $\hat{K}_S = 0.572$  mg/L, and  $\hat{Y} = 0.413$  mg/mg. For 2MN, the ratio of  $\hat{q}_{\max}/\hat{K}_S$  was estimated to be  $0.193$  mg/mg h<sup>-1</sup>, with  $\hat{Y} = 0.381$  mg/mg.

Comparison of the observations to the model-predicted results demonstrates a good fit to both the substrate and biomass data for NPH (Fig. 1). The substrate data range over five orders of magnitude and the different regions of the Monod equations are demonstrated. The substrate depletion is initially near the zero-order rate regime, because  $C > K_S$ , and then passes through the mixed-order regime and into the first-order regime when  $C \ll K_S$ . The presence of two distinct linear regimes is characteristic of the Monod curve in log space. The 2MN fit is not as good as for NPH. The substrate depletion experimental data seem to demonstrate an S-shape in log space. The Monod formulation is unable to capture the decrease in substrate depletion rate after long time periods. In addition, biomass growth is not completely captured. It is unclear whether there is a lag in the biomass growth or if the error in the biomass measure-



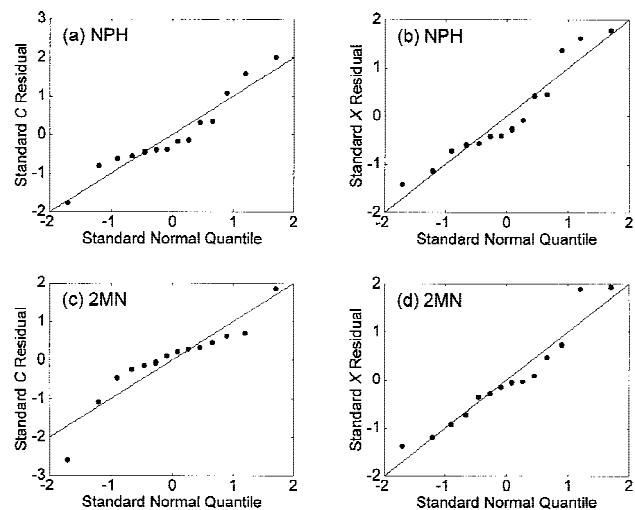
**Figure 1.** Comparison of model fit (solid curves) to observations (symbols).

ments was so large that the model is capturing the biomass growth within the range of reasonable error. The implications of this lack of fit are discussed later.

## Residual Analyses

The assumptions made in the derivation of the maximum likelihood formulation were checked for validity through analyses of the residuals. The assumptions were that the errors have a normal distribution with a mean of zero, have constant variance, and are independent within each variable type.

A normal probability plot of residuals is linear if the measured variables have normally distributed errors (Mason et al., 1989). The parameter estimation algorithm was first performed on the untransformed substrate and biomass data (i.e.,  $Y_1 = C$  and  $Y_2 = X$ ) (Fig. 2). A linear result was



**Figure 2.** Quantile-normal quantile plots of residuals for the regression of the untransformed  $C$  and  $X$  data.

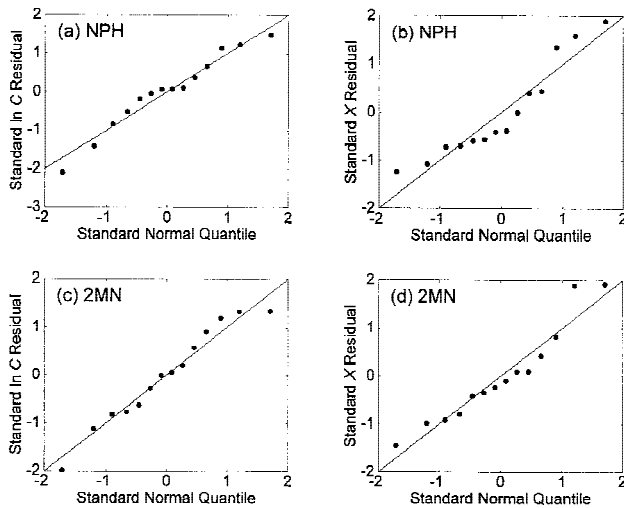
demonstrated for biomass residuals, suggesting that the errors are indeed normal. The residuals for substrate concentration, however, were not linear with the  $y = x$  line. This suggests that errors in substrate concentration were not described by a normal distribution. The parameter estimation was rerun on data in which the substrate concentration had been transformed into log-normal space. When these residuals were examined, the residuals for  $\ln C$  fell more appropriately upon the  $y = x$  line (Fig. 3). This suggests that the errors in substrate concentration were distributed log-normally.

If the underlying errors have asymmetric distributions, there is a possibility that the mean of the errors is nonzero. A statistical test of the residuals determines if the mean of the residuals is significantly different than zero. Defining  $\bar{e}$  as the average residual vector, and  $\hat{e}_j$  as the estimated residual vector [Eq. (7)], the statistic:

$$\lambda = \left[ \frac{(n-m)n}{(n-1)m} \right] \bar{e}^T \left[ \frac{1}{n-1} \sum_{i=1}^n (\hat{e}_i - \bar{e})(\hat{e}_i - \bar{e})^T \right]^{-1} \bar{e} \quad (19)$$

is distributed as  $F_{m,n-m}$  (Bard, 1974). For the two substrates,  $\lambda_{\text{NPH}} = 0.214$  and  $\lambda_{2\text{MN}} = 0.472$ . These values correspond to confidence levels of 88.5% and 70.8%, respectively. This suggests that there was relatively high confidence that the errors have a mean of zero. This further validates the necessity for transforming substrate concentration into log space. When the regression was performed on the untransformed substrate concentration, the test yielded  $\lambda = 1.60$  for NPH and  $\lambda = 3.60$  for 2MN, corresponding to confidence levels of only 24% and 5.7%, respectively.

The residuals were analyzed for nonrandomness by plotting the residuals versus the estimated values and versus time. For NPH, the substrate and biomass data shown in Figure 4 a–d demonstrate random scatter, suggesting that the errors are independent and have a constant variance. Because each observation was made from an identical but



**Figure 3.** Quantile-normal quantile plots of residuals for regression of log-transformed  $C$  data and untransformed  $X$  data.

distinct experiment, it is reasonable to believe that the approach used would result in data that are not serially correlated. The serial correlation was checked quantitatively by calculating the number of runs of residuals, and then calculating the probability that these runs would occur randomly. The numbers of negative residuals ( $n_1$ ) and positive residuals ( $n_2$ ), and number of runs ( $u$ ) of residuals were counted. These values were evaluated using the cumulative distribution functions for number of runs with  $n_1$  and  $n_2 \leq 10$  (Draper and Smith, 1981). The results in Table I demonstrate that the substrate concentration data are random with good confidence. The confidence is low for the number of runs occurring randomly for the biomass data. This is due to the large run of positive residuals in the intermediate time of the experiment.

For 2MN, there seems to have been a lack of randomness in the residuals for both measured variables (Fig. 4e–h). For the errors in substrate concentration, the residuals were generally positive when the concentration was high and negative when the concentration was low. This corresponded to a series of positive residuals in the intermediate times and negative residuals at later times. The biomass residuals showed similar problems of being negative at early to intermediate times, and positive at later times. When the test for the runs of residuals was performed a low probability of randomness was demonstrated for both variables. These plots and the test of the residuals suggest a possible inadequacy in the model. This statistical analysis demonstrates that the Monod model was unable to capture both the subtle S-shape of the substrate data and the extent of biomass growth in the 2MN experiments, which was mentioned earlier from visual inspection of measured data versus model fit (Fig. 1).

### Sample Covariance Matrix

The covariance matrix of the errors is written as:

$$\hat{\Sigma} = \begin{bmatrix} \sigma_{\ln C}^2 & \sigma_{\ln C, X}^2 \\ \sigma_{\ln C, X}^2 & \sigma_X^2 \end{bmatrix} \quad (20)$$

and these were estimated with the parameters during the optimization procedure [see Eq. (14)]. The estimates were:

$$\hat{\Sigma}_{\text{NPH}} = \begin{bmatrix} 0.076 & 0.037 \\ 0.037 & 1.12 \end{bmatrix} \quad \hat{\Sigma}_{2\text{MN}} = \begin{bmatrix} 0.144 & -0.316 \\ -0.316 & 2.97 \end{bmatrix}$$

Because  $C$  was estimated in log space, the variance must be transformed into absolute space. When the relative error is small ( $<10\%$ ), the variance in  $C$  can be approximated from the relation:

$$\hat{\sigma}_{\ln C}^2 \cong \frac{\hat{\sigma}_C^2}{C^2} \quad (21)$$

The square root of this value is the estimated relative standard error in  $C$ . The standard errors for the variables  $C$  and  $X$  are given in Table II.

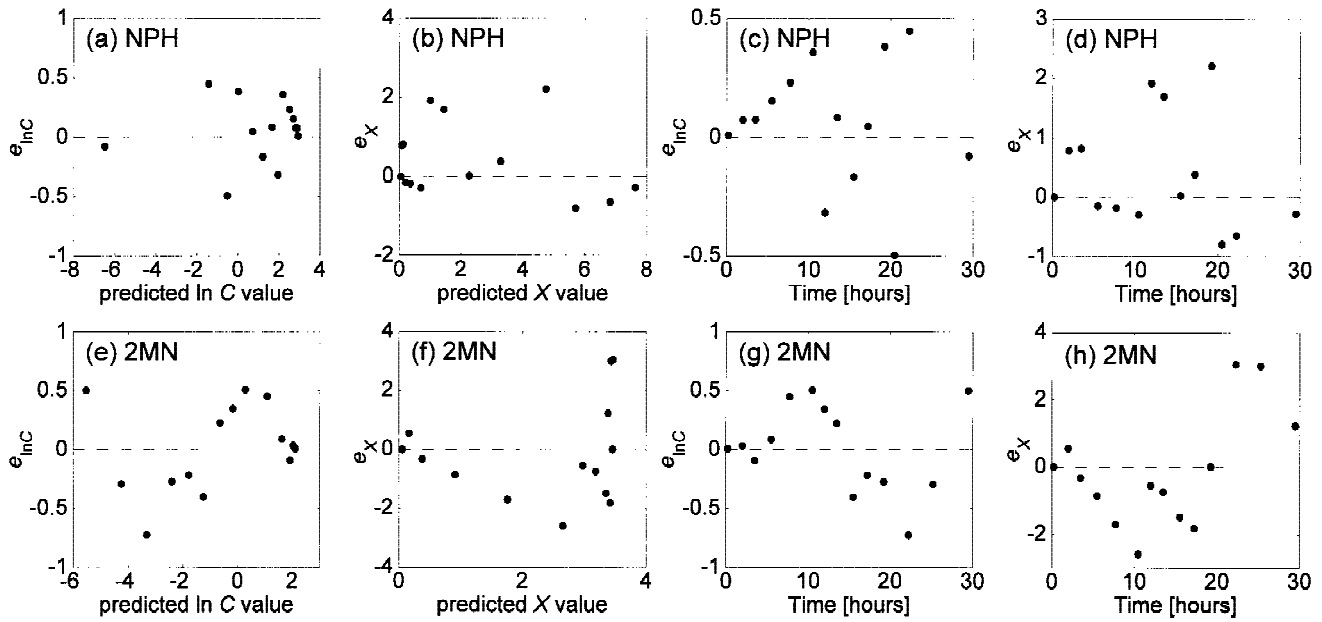


Figure 4. Residuals vs. predicted values of regression variables and residuals vs. time.

The results of the covariance matrices demonstrate that there are indeed different error variances for  $\ln C$  and  $X$ . In both cases, the variance of  $\ln C$  is much less than the variance of  $X$ . If these were assumed to be the same, then the parameter estimation would have put too much weight on the biomass values. Also, the values of the covariance matrices are different for NPH and 2MN. Therefore, there is no basis a priori for assuming values of the weights of each variance. This is one reason for using the more extensive maximum likelihood estimation method with an unknown covariance matrix over simpler methods.

The covariance,  $\hat{\sigma}_{\ln C, X}^2$  is small but not negligible. To determine the significance of the covariance, the correlation coefficient,  $r_{\ln C, X}$  is calculated:

$$r_{\ln C, X} = \frac{\hat{\sigma}_{\ln C, X}^2}{(\hat{\sigma}_{\ln C}^2 \hat{\sigma}_X^2)^{1/2}} \quad (22)$$

Using  $r_{\ln C, X}$ , the statistic:

$$\lambda = r_{\ln C, X} \left( \frac{n^* - 2}{1 - r_{\ln C, X}^2} \right)^{1/2} \quad (23)$$

has the  $t$ -distribution with  $n^* - 2$  degrees of freedom, where  $n^* = n - l/m$  (Bard, 1974). For these analyses,  $n^* = 12$  for

**Table I.** The probability of the number of runs occurring randomly ( $n_1$  is number of negative residuals,  $n_2$  is number of positive residuals,  $u$  is number of expected runs).

|                 | $n_1$ | $n_2$ | $u$ | Probability |
|-----------------|-------|-------|-----|-------------|
| NPH ( $\ln C$ ) | 4     | 10    | 8   | 87.4%       |
| NPH ( $X$ )     | 7     | 7     | 5   | 7.8%        |
| 2MN ( $\ln C$ ) | 6     | 10    | 5   | 4.7%        |
| 2MN ( $X$ )     | 10    | 4     | 4   | 6.8%        |

both NPH and 2MN. For NPH,  $r_{\ln C, X} = 0.127$  and  $\lambda = 0.452$  and, for 2MN,  $r_{\ln C, X} = -0.485$  and  $\lambda = -1.96$ . For 2MN, the probability that  $|t_{10}| > 1.96$  is 6%, which gives a 94% confidence that  $\sigma_{\ln C, X}^2 \neq 0$ . For NPH, the probability that  $|t_{10}| > 0.452$  is 33%, so there is some, albeit limited, justification for assuming that the covariance is nonzero. The nonzero covariance further demonstrates the need for using maximum likelihood estimation over weighted least squares regression.

### Confidence Limits

The confidence in the parameter estimation was determined by estimating the covariance matrix of the estimates:

$$V_{\hat{\theta}} = \begin{bmatrix} \sigma_{\ln q}^2 & \sigma_{\ln q, \ln K_S}^2 & \sigma_{\ln q, Y}^2 \\ \sigma_{\ln q, \ln K_S}^2 & \sigma_{\ln K_S}^2 & \sigma_{\ln K_S, Y}^2 \\ \sigma_{\ln q, Y}^2 & \sigma_{\ln K_S, Y}^2 & \sigma_Y^2 \end{bmatrix} \quad (24)$$

Assuming that the errors in the parameters are distributed normally, the covariance matrix of the estimates can be estimated by:

$$V_{\hat{\theta}} \approx \hat{H}^{-1} = - \left( \frac{\partial^2 \log L}{\partial \underline{\theta}^2} \right)_{\hat{\theta}}^{-1} \quad (25)$$

**Table II.** Estimated standard errors for substrate concentration data (relative errors) and for biomass concentration data (absolute errors).

| Variable type | Estimated standard error ( $\hat{\sigma}$ ) |
|---------------|---|
| $C$ , NPH     | 28%   |
| $X$ , NPH     | 1.06  |
| $C$ , 2MN     | 38%   |
| $X$ , 2MN     | 1.72  |

where  $\hat{H}$  is the Hessian evaluated at  $\underline{\theta} = \hat{\underline{\theta}}$ . Because  $L$  and  $\Phi$  differ by only a constant, the Hessian can be written as :

$$\underline{H}(\underline{\theta}) = -\frac{\partial^2 \log L}{\partial \underline{\theta}^2} = \frac{\partial^2 \Phi}{\partial \underline{\theta}^2} \quad (26)$$

If the Hessian cannot be determined analytically by differentiation, it is often approximated using the Gauss method, which neglects all the second derivatives of the model equations (Sleep and Mulcahy, 1998; Smith et al., 1997). If the residuals are very small, then this approximation may be acceptable. Instead, an alternative method for determining the Hessian was performed by numerically approximating the objective function about  $\underline{\theta}$ . Three grids of  $42 \times 42$  points were formed on orthogonal planes by varying two parameters and holding the third at its optimal value. A response surface was generated in each plane by computing the value of  $\Phi$  at each grid point. Points were selected whose  $\Phi$  values fell within the 95% confidence region. An estimate for the value of the objective function corresponding to a given level of confidence  $(1 - \alpha)$  was obtained by computing:

$$\Phi(\hat{\underline{\theta}}) \left\{ 1 + \frac{l}{mn-l} F(l, mn-l, 1-\alpha) \right\} \quad (27)$$

In the case of nonlinear parameter estimation, the probability level is approximate (Draper and Smith, 1986).

A second-order polynomial [Eq. (28)] was fit using a response surface technique (Box et al., 1978) to approximate the objective function:

$$\begin{aligned} \Phi(\ln q_{\max}, \ln K_S, Y) = & \beta_0 + \beta_1(\ln q_{\max}) + \beta_2(\ln K_S) + \beta_3 Y \\ & + \beta_4(\ln q_{\max})(\ln K_S) + \beta_5(\ln q_{\max})Y \\ & + \beta_6(\ln K_S)Y + \beta_7(\ln q_{\max})^2 \\ & + \beta_8(\ln K_S)^2 + \beta_9 Y^2 \end{aligned} \quad (28)$$

Visual comparisons of contour plots of the resulting polynomial and the  $\Phi$  response surfaces indicate that the approximation was excellent. The resulting polynomial was differentiated to estimate the Hessian. Therefore, with the power and speed of using computers for numerical calculations, the Gaussian method of approximating the Hessian was not necessary.

The covariance matrix for NPH is:

$$\underline{V}_{\underline{\theta}, \text{NPH}} = \begin{bmatrix} 1.01 \times 10^{-4} & 2.70 \times 10^{-5} & -3.86 \times 10^{-4} \\ 2.70 \times 10^{-5} & 3.27 \times 10^{-6} & 6.18 \times 10^{-6} \\ -3.86 \times 10^{-4} & 6.18 \times 10^{-6} & 1.41 \times 10^{-3} \end{bmatrix}$$

From the diagonal elements of this matrix, the 95% confidence bounds were determined (Table III). For  $Y$ , this was estimated as  $Y \pm t_{n^*, \frac{\alpha}{2}} \hat{\sigma}_Y$ , using  $n^* = 12$ ,  $t_{0.025} = 2.18$ . For  $q_{\max}$  and  $K_S$ , these were expressed as relative errors [Eq. (21)] scaled by the same  $t$ -statistic. The 95% confidence bounds were on the order of 2% for  $q_{\max}$ , 0.4% for  $K_S$  and 20% for  $Y$ . Inspection of the covariance matrix of the estimates demonstrated a certain degree of correlation. There-

**Table III.** Parameter results and their relative errors (at 95% confidence) for those parameters for which the covariance matrix was estimated in log space, or 95% confidence.

| Parameter                              | NPH           | 2MN           |
|--|---------------|---------------|
| $q_{\max}$ (mg/mg h <sup>-1</sup> )    | 0.636 (2%)    | n/a           |
| $K_S$ (mg/L)                           | 0.572 (0.4%)  | n/a           |
| $q_{\max}/K_S$ (mg/L h <sup>-1</sup> ) | 1.11 (2%)     | 0.193 (4%)    |
| $Y$ (mg/mg)                            | 0.413 ± 0.082 | 0.381 ± 0.089 |

fore, a more thorough understanding of the confidence in the estimates was obtained by estimating confidence regions in a two-dimensional parameter space.

An approximation for the contour encompassing each joint confidence region was given by plotting the response surface polynomial [Eq. (28)] at a series of different confidence contours [Eq. (27)]. For all combinations of parameters, there was a certain amount of correlation between all of the parameters (Fig. 5). Correlation is an inherent problem with a model such as the Monod equation, which has regimes described by simpler functional forms with fewer parameters. Because of the appreciable correlation, the confidence bounds of the individual parameters were only estimates.

Comparing the joint confidence regions of NPH and 2MN, it is evident that there was more confidence in the parameters determined for NPH. In addition, it is evident that there was a consistency in the sign of correlation. There was positive correlation between  $q_{\max}$  and  $K_S$ , and  $Y$  and  $K_S$ , and a negative correlation between  $q_{\max}$  and  $Y$ . The correlation of parameters with  $Y$  was more pronounced for 2MN than it was for NPH. This was demonstrated by the larger size that the 95% confidence region encompassed for each pair of parameters.

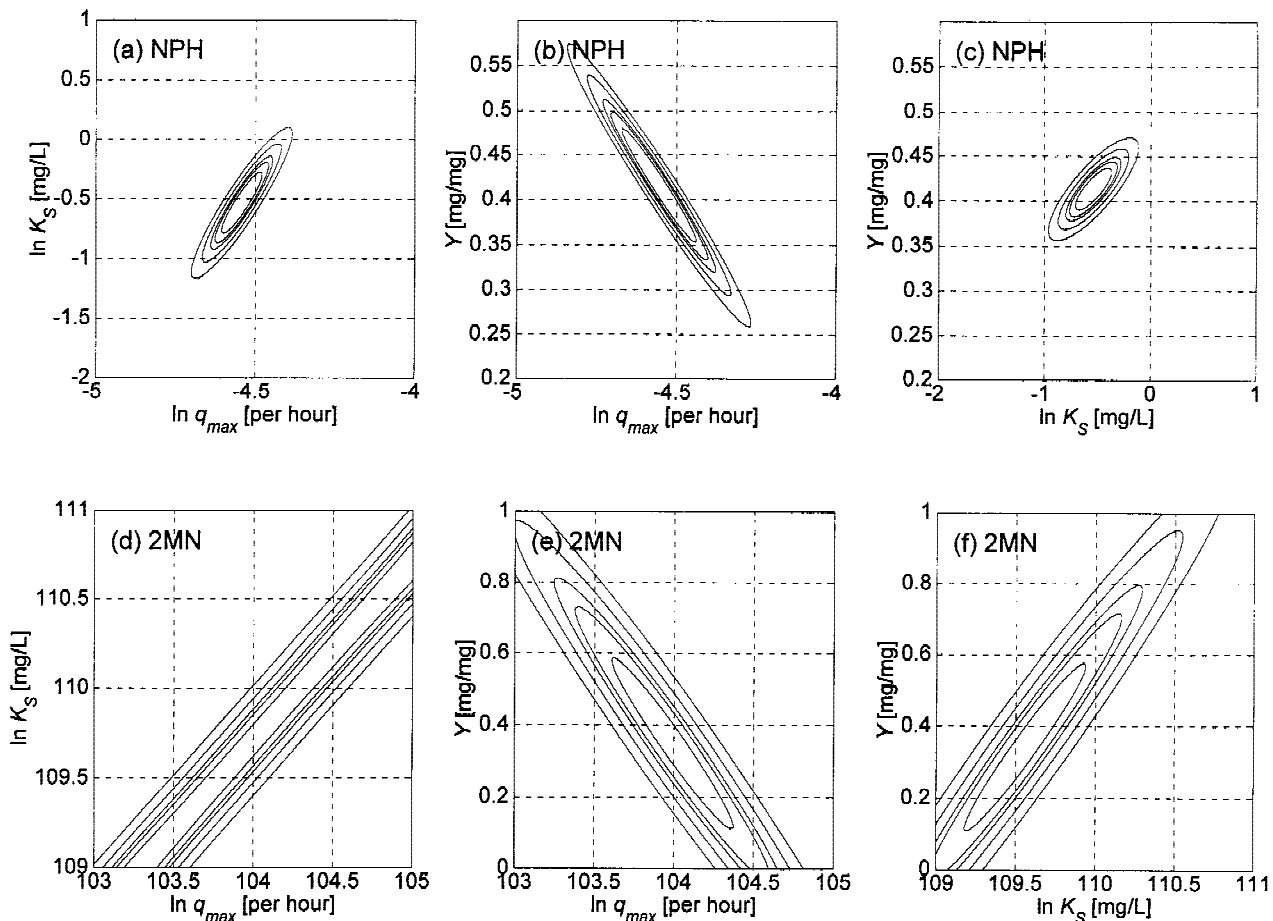
The joint confidence regions for 2MN are presented in Figure 5d-f. The optimization process revealed the lack of a unique minimum of the objective function in the  $q_{\max}$ - $K_S$  space. Instead, the objective function had increasingly smaller values heading toward infinity with a constant ratio of  $q_{\max}:K_S$ . The correlation between  $q_{\max}$  and  $K_S$  is evident in the approximations of the joint confidence regions of  $q_{\max}$  and  $K_S$  presented in Figure 5d as a well-defined valley. The variance of  $q_{\max}/K_S$  was determined using:

$$\frac{\sigma_k^2}{k^2} = \frac{\sigma_{q_{\max}}^2}{q_{\max}^2} + \frac{\sigma_{K_S}^2}{K_S^2} - 2 \frac{\sigma_{q_{\max}K_S}}{q_{\max}K_S} \quad \text{where: } k = \frac{q_{\max}}{K_S} \quad (29)$$

from Bevington and Robinson (1992). Using the relationship in Eq. (21),  $\sigma_k^2$  was estimated. The first-order biodegradation rate for 2MN was estimated as  $q_{\max}/K_S = 0.193$  L/mg h<sup>-1</sup>, with a 95% confidence bound of 4%.

### Comparison with Univariate Approach

Measurement of biomass is difficult and the resulting data have large uncertainties. Here, we examine the extent to



**Figure 5.** Joint confidence regions for the estimated parameters (from innermost to outermost: 75%, 90%, 95%, 99%, 99.9%). (a) NPH: contours in the  $\ln q_{\max}$ - $\ln K_S$  plane at  $Y = \hat{Y}$ ; (b) NPH: contours in the  $\ln q_{\max}$ - $Y$  plane at  $K_S = \hat{K}_S$ ; (c) NPH: contours in the  $\ln K_S$ - $Y$  plane at  $q_{\max} = \hat{q}_{\max}$ ; (d) 2MN: contours in the  $\ln q_{\max}$ - $\ln K_S$  plane at  $Y = \hat{Y}$ ; (e) 2MN: contours in the  $\ln q_{\max}$ - $Y$  plane at  $K_S = \hat{K}_S$ ; (f) 2MN: contours in the  $\ln K_S$ - $Y$  plane at  $q_{\max} = \hat{q}_{\max}$ .

which biomass measurement data improve the parameter estimation. A nonlinear parameter estimation was performed using only the substrate concentration data and the initial biomass concentration. If only substrate concentrations had been measured, as is commonly done, then the maximum likelihood method would reduce to the familiar nonlinear least squares regression technique. The objective function for the univariate nonlinear least squares regression is given in Eq. (10). The same optimization technique was used for the univariate case as outlined for the bivariate case.

The resulting estimated parameters with single confidence limits are listed in Table IV. Even though the same data set of substrate concentrations was used in the parameter estimations, different values were obtained for the parameters and their confidence limits. The precision of the estimates was altered when the biomass data were not measured and incorporated into the parameter estimation. For NPH, the precision decreased, as one would expect when fewer data are included. For 2MN, however, the precision appeared to increase, which suggests that the errors in the biomass data were so large as to diminish the precision of the estimates. A different picture emerges if we consider the

accuracy of the estimates. If the parameters with their 95% confidence bounds are overlaid, they do not overlap. The inconsistency between the two estimations implies that the biomass data may provide unique information.

By examining the univariate modeling fits compared with both data sets it is evident that the biomass data fit poorly (not shown), which suggests a lack of consistency when the biomass data are not included. For NPH, the biomass growth was dramatically overestimated and  $Y$  was quite

**Table IV.** Parameter results and their relative errors (at 95% confidence) for those parameters for which the covariance matrix was estimated in log space, or 95% confidence limits as  $\pm$  using maximum likelihood estimation (MLE) using substrate and biomass concentration data and nonlinear least squares (NLS) using only substrate concentration data.

| Compound parameter                         | MLE value         | NLS value         |
|--|-------------------|-------------------|
| NPH $q_{\max}$ (mg/mg h <sup>-1</sup> )    | 0.636 (2%)        | 0.346 (4%)        |
| NPH $K_S$ (mg/L)                           | 0.572 (0.4%)      | 0.291 (1%)        |
| NPH $q_{\max}/K_S$ (mg/L h <sup>-1</sup> ) | 1.11 (2%)         | 1.19 (1%)         |
| NPH $Y$ (mg/mg)                            | 0.413 $\pm$ 0.082 | 0.775 $\pm$ 0.159 |
| 2MN $q_{\max}/K_S$ (mg/L h <sup>-1</sup> ) | 0.193 (4%)        | 0.267 (2.5%)      |
| 2MN $Y$ (mg/mg)                            | 0.381 $\pm$ 0.089 | 0.241 (8%)        |

large ( $Y = 0.775$ ). For 2MN, the reverse occurred and the biomass growth was underestimated ( $Y = 0.241$ ). Because of the impact of biomass growth on the biodegradation kinetics, and due to the sorption capacity of the biomass, the biodegradation rate parameters were altered accordingly.

From the bivariate regression, the model fit the NPH data well and all residual tests validated assumptions. For NPH, the additional biomass information increased the precision of the estimates. As previously mentioned, however, the 2MN data were not fit as well by the predicted parameters in the bivariate case. This lack of fit manifested itself in the residual analysis as an implication of residual correlation. The S-shape curve in the substrate concentration data plotted in log space was revealed by the univariate approach, but the difficulty in modeling biomass growth cannot be understood. The difficulty in modeling the biomass curve would have been lost, however, and very different values for  $Y$  would have been estimated. Furthermore, a false sense of heightened precision would have been inferred for the 2MN parameters if biomass was not measured.

The lack of the biomass data removes a vital piece of information from the estimation procedure, despite its high uncertainty. The inability to accurately model the biomass growth curve without incorporating the biomass data in the parameter estimation exemplifies the importance of measuring the biomass concentration, its use in the parameter estimation method, and therefore the necessity for using the nonlinear parameter estimation technique outlined here.

## CONCLUSIONS

The maximum likelihood estimation approach was used to develop a method for estimating the Monod parameters  $q_{\max}$ ,  $K_S$ , and  $Y$ , in the case where measurements have been taken for two variables. The method avoids unnecessary simplifying assumptions regarding normality of errors, constant error variance, and the lack of covariance between measured variables. Analysis of regression residuals revealed that the errors in the substrate concentration measurements did not have a normal distribution, whereas the biomass concentrations did. Transformation of substrate concentrations into logarithmic space resulted in regression statistics that validated the assumptions inherent in the derivation of the parameter estimation objective function. The estimated covariance matrix of the variables was nondiagonal, indicating significant nonzero covariance. The biomass concentrations had a greater variance than the log-transformed substrate concentrations, indicating a lack of constancy of the errors across the two data types. These findings validate the need to use maximum likelihood estimation rather than a simpler, bivariate least squares regression approach.

Response surface analysis of the parameter estimation objective function over the  $q_{\max}$ - $K_S$  space revealed that a true minimum exists for the NPH data set, but not for the 2MN data set. Thus, unique values of  $q_{\max}$ ,  $K_S$ , and  $Y$  were estimated for NPH, although there was some correlation

between the parameters. For 2MN, the first-order rate parameter,  $q_{\max}/K_S$ , was estimated. For this substrate, evidence of serial correlation in the substrate concentration residuals revealed phenomena that the Monod model cannot adequately capture. For both substrates, it was found that a better approximation of the confidence limits and joint confidence regions of the parameters was obtained after a logarithmic transformation of the  $q_{\max}$  and  $K_S$  parameters.

The larger variance in the biomass concentration errors, relative to the more precise substrate concentration measurements, leads to the question of whether these data facilitate the parameter estimation task. Our comparison of the maximum likelihood estimation approach using the bivariate data sets and a nonlinear least squares regression technique using only substrate concentration data revealed that the univariate case resulted in a poor fit for biomass data. By using both measured variables the inability of the Monod formulation to completely capture the biomass growth in the 2MN experiment was realized. Thus, despite the large uncertainties in biomass measurements, the inclusion of different data types leads to insights regarding inconsistencies between measured variables and possible inadequacies in model formulation.

The maximum likelihood method is a more rigorous approach than nonlinear least squares and it requires a more complicated computer algorithm for execution. It is, however, not only more thorough but also necessary when using a bivariate data set with different error structures. The additional information gained when measuring concentration as well as biomass, despite the inherent large variance of biomass, is valuable when estimating Monod parameters.

## NOMENCLATURE

|                       |   |
|-----------------------|---|
| 2MN                   | 2-methylnaphthalene   |
| $b$                   | endogenous decay rate coefficient ( $\text{h}^{-1}$ )                                     |
| $C$                   | concentration of the substrate (mg/L)   |
| $C_0$                 | initial concentration of the substrate (mg/L)   |
| $\underline{e}$       | vector of the residuals   |
| $i$                   | given observation   |
| $F$                   | $F$ -distribution   |
| $\underline{H}$       | Hessian   |
| $k$                   | first-order biodegradation rate (L/mg [protein]/h)  |
| $k_a$                 | abiotic loss rate coefficient ( $\text{h}^{-1}$ )   |
| $K_a$                 | abiotic partition coefficient (dimensionless)   |
| $K_b$                 | coefficient for partitioning to biorelated material (L/mg [protein])                      |
| $K_S$                 | half-saturation coefficient (mg/L)  |
| $L$                   | likelihood function   |
| $l$                   | number of estimated parameters  |
| $\underline{M}$       | moment matrix   |
| $m$                   | number of variables   |
| $n$                   | number of observations  |
| $n_1$                 | number of negative residuals  |
| $n_2$                 | number of positive residuals  |
| NPH                   | naphthalene   |
| PAH                   | polycyclic aromatic hydrocarbon   |
| $q_{\max}$            | maximum substrate utilization rate per unit biomass (mg substrate/mg biomass [protein]/h) |
| $r_{\text{in } C, X}$ | correlation coefficient for the variables in $C$ and $X$                                  |
| $Tr$                  | trace of a given matrix   |

|            |   |
|------------|---|
| $T$        | transpose of a given matrix                           |
| $t$        | time (h)  |
| $u$        | number of runs of residuals                           |
| $V$        | covariance matrix                                     |
| $\bar{X}$  | concentration of the biomass (mg [protein]/L)         |
| $X_0$      | initial concentration of the biomass (mg [protein]/L) |
| $Y$        | yield coefficient (mg [protein]/mg)                   |
| $Y_1, Y_2$ | measured variable 1 or 2                              |

#### Greek symbols

|            |  |
|------------|--|
| $\alpha$   | probability level  |
| $\beta$    | coefficient for polynomial approximation of response surface |
| $\lambda$  | statistic used for significance testing                      |
| $\sigma^2$ | covariance   |
| $\theta$   | vector of parameters   |
| $\Phi$     | objective function   |

#### Special symbols

|          |                     |
|----------|---------------------|
| $\wedge$ | Estimated parameter |
| $\sim$   | vector or matrix    |

## References

- Bard Y. 1974. Nonlinear parameter estimation. New York: Academic Press.
- Bevington PR, Robinson DK. 1992. Data reduction and error analysis for the physical sciences. New York: McGraw Hill.
- Box GEP, Hunter WG, Hunter JS. 1978. Statistics for experimenters: an introduction to design, data analysis, and model building. New York: John Wiley & Sons.
- Draper NR, Smith H. 1981. Applied regression analysis. New York: John Wiley & Sons.
- Ellis TG, Barbeau DS, Smets CP, Grady L Jr. 1996. Respirometric technique for determining extant kinetic parameters describing biodegradation. *Water Environ Res* 68:917–926.
- Guha S, Jaffe PR. 1996. Determination of Monod kinetic coefficients for volatile hydrophobic organic compounds. *Biotechnol Bioeng* 50:693–699.
- Knights CD. 2000. Mechanisms governing sole-substrate and multisubstrate biodegradation kinetics of polycyclic aromatic hydrocarbons. PhD dissertation. Princeton University, Princeton, NJ.
- Leatherbarrow RJ. 1990. Using linear and non-linear regression to fit biochemical data. *Trends Biochem Sci* 15:455–458.
- Mason RL, Gunst RF, Hess JL. 1989. Statistical design and analysis of experiments with applications to engineering and science. New York: John Wiley & Sons.
- Ong SL. 1983. Least squares estimation of batch culture kinetic parameters. *Biotechnol Bioeng* 25:2347–2358.
- Robinson JA. 1985. Determining microbial kinetic parameters using nonlinear regression analysis. *Adv Microbial Ecol* 8:61–114.
- Robinson JA. 1998. Modeling microbial processes: an overview of statistical considerations. In: Koch AL, Robinson JA, Milliken GA, editors. *Mathematical modeling in microbial ecology*. New York: Chapman & Hall.
- Robinson JA, Tiedje JM. 1983. Nonlinear estimation of Monod growth kinetic parameters from a single substrate depletion curve. *Appl Environ Microbiol* 45:1453–1458.
- Sleep BE, Mulcahy LJ. 1998. Estimation of biokinetic parameters for unsaturated soils. *J Environ Eng* 124:959–969.
- Smith L, Kitanidis PK, McCarty PL. 1997. Numerical modeling and uncertainties in rate coefficients for methane utilization and TCE cometabolism by a methane-oxidizing mixed culture. *Biotechnol Bioeng* 53:320–331.
- Smith LH, McCarty PL, Kitanidis PK. 1998. Spreadsheet method for evaluation of biochemical reaction rate coefficients and their uncertainties by weighted nonlinear least squares analysis of the integrated Monod equation. *Appl Environ Microbiol* 64:2044–2050.
- Sommer HM, Holst H, Spliid H, Arvin E. 1995. Nonlinear parameter estimation in microbiological degradation systems and statistic test for common estimation. *Environ Int* 21:551–556.