Estimation of Parameters Where Dependent Observations Are Related by Equality Constraints

Pirt's model for microbial growth and product formation are reparameterized to obtain multiresponse models with common parameters. The dependent variables in the models are related through the available electron and carbon balance constraints. Covariance adjustment is used to reduce the growth model to a unit variate linear model with covariates. Therefore, standard multiple regression programs can be used to obtain combined point and interval estimates of true biomass energetic yield, true product yield and maintenance coefficient. This approach may yield "better" estimates than the maximum likelihood approach when an appropriately selected subset of covariates is used. Nonlinear estimation procedures are also considered; these procedures are efficient with few responses; however, as the number of responses per observation increase, they may require a lot of computing time. For illustration several data from the biochemical engineering literature are analyzed by the proposed methods.

B. O. SOLOMON, M. D. ONER, and L. E. ERICKSON

Department of Chemical Engineering
Kansas State University
Manhattan, KS 66502

S. S. YANG

Department of Statistics
Kansas State University
Manhattan, KS 66502

SCOPE

Advances in genetic engineering are being applied by microbiologists, plant scientists and animal scientists to develop new and improved strains of a wide variety of growing organisms. In most cases, efficient growth is an important objective. In other cases, product formation efficiency is important. For each improved strain which is developed, optimum conditions for growth and product formation are also of interest.

This work is concerned with the application of statistical methods of parameter estimation, and data consistency analysis to growth and product formation where the dependent variables are related by equality constraints. In earlier work, parameter estimation results with point and interval estimates are reported by Erickson and Hess (1981), and by Solomon et al. (1981, 1982). Data consistency analysis and methodology to use all of the available measurements to obtain maximum likelihood estimates have also been reported previously by Solomon et al. (1982).

When more than the minimum number of variables are measured, if measurement error is considered, parameter estimation results depend on which of the measured variables are selected for use in parameter estimation. Models relating the response with the common parameters can be put into a type of multivariate linear model which is a special case of the general growth curve model introduced by Putthoff and Roy (1964). In earlier work (Solomon et al., 1982), this type of model was used to analyze data obtained from fermentation processes. The method of analysis was based on results obtained by Khatri (1966) and Rao (1967). In this paper, using Rao's (1965, 1966, 1967) ideas of covariance adjustment and following Grizzle and Allen (1969), this special type of growth curve model can be reduced to a unit variate linear model with covariates. Therefore, standard multiple regression programs can be used to estimate parameters and to test hypotheses concerning the parameters. Moreover, the covariance adjustment approach provides additional flexibility in weighting by choosing subsets of covariates. An approximately selected subset of covariates may in some cases yield "better" estimates than using all the covariates. In this work, the results show that the estimates with shorter 95% confidence intervals are frequently obtained when covariate adjustment is used.

Also presented are some nonlinear parameter estimation techniques using the Hooke and Jeeves (1961) Pattern Search Technique to estimate parameters by minimization of some determinant criteria which have been considered previously by Oner et al. (1983) and Box and Draper (1965). The 95% confidence intervals of the parameters are constructed using the Smooth Bootstrap method (Efron, 1979, 1981).

The applicability of the linear as opposed to the non-linear techniques are discussed from a computational point of view.

CONCLUSION AND SIGNIFICANCE

Application of the available electron and carbon balances and the associated regularities has unified all the different types of yields and maintenance coefficients in Eqs. 16–19; more than one estimate of the yield and maintenance parameters may be obtained. The statistical methodologies which are presented allow one to combine the different estimates and has led to improved point and interval estimates for the true biomass energetic yield, true product yield and maintenance coefficients.

The maximum likelihood estimator using all the covariates is not always the best estimator. In fact including only appropriately selected subsets of covariates sometimes yields better estimators. The "dominant characteristics vector" method is also an efficient and simple method for determining useful covariates.

The results also show that the parameter estimates are sensitive to the form of the equations. Forms I and II utilized with the product formation case give results which differ some and utilization of both forms is desirable because this gives a clear indication of the dependence of the estimates on the form of the equations.

The growth curve model is a very general model. It is easily applied to estimate yields and maintenance parameters associated with growth and product formation. The techniques considered in this article will also be useful to analyze multiresponse data from other contexts.

INTRODUCTION

Aerobic microbial processes as well as other aerobic processes such as those associated with animal and human growth and nutrition may be analyzed using available electron and energetic yield concepts described by Minkevich and Eroshin (1973), Payne (1970), and Erickson et al. (1978). Recently, Erickson and Patel (1981, 1982) utilized these concepts in an examination of energetic yields and efficiencies of a wide variety of growth processes.

An available electron balance or energy balance based on the chemical energy in the organic substrate (food) utilized by the growing organism may be written as

$$\begin{matrix}\text{Energy}\\ \text{Input}\end{matrix} = \begin{matrix}\text{Energy allocated}\\ \text{for biomass}\\ \text{production}\end{matrix} + \begin{matrix}\text{Maintenance}\\ \text{energy}\end{matrix} + \begin{matrix}\text{Energy allocated}\\ \text{for product}\\ \text{formation}\end{matrix} \tag{1}$$

Based on the chemical energy in the organic substrate, the energetic efficiency of growth is (Erickson, 1979)

$$\eta_{\max} = \frac{\text{energy incorporated in the biomass}}{\text{energy allocated for biomass production}} \tag{2}$$

Similarly, the energetic efficiency of product formation is

$$\xi_p^{\max} = \frac{\text{energy incorporated into products}}{\text{energy allocated for product formation}} \tag{3}$$

Another form of the energy balance may also be written. Based on the final products,

$$\begin{matrix}\text{Energy}\\ \text{Input}\end{matrix} = \begin{matrix}\text{Energy incorporated}\\ \text{into biomass}\end{matrix} + \begin{matrix}\text{Energy lost}\\ \text{as heat}\end{matrix} + \begin{matrix}\text{Energy incorporated}\\ \text{into products}\end{matrix} \tag{4}$$

or dividing by the energy input

$$1 = \eta + \epsilon + \xi_p \tag{5}$$

where η is the biomass energetic yield, ϵ is the fraction of consumed energy evolved as heat (or the fraction of available electrons transferred to oxygen), and ξ_p is the product energetic yield. Equation 1 may be divided by the energy input and written in the form

$$1 = \frac{\eta}{\eta_{\max}} + \epsilon_m + \frac{\xi_p}{\xi_p^{\max}} \tag{6}$$

where ϵ_m is the fraction of input energy evolved as heat because of maintenance.

The parameters $\eta_{\max}$ and $\xi_p^{\max}$ are of considerable commercial importance, and thus, they are to be estimated. The quantities η, ϵ and ξ_p may be calculated from appropriate measurements. The quantities, η, ϵ, and ξ_p depend on growth rate while $\eta_{\max}$ and $\xi_p^{\max}$ are assumed to be independent of growth rate. In addition to these two parameters, a maintenance parameter is also estimated; however, the form of parameter depends on the species being considered. For microbial growth, where μ is the specific growth rate, Eq. 6 may be written in the form (Erickson, 1979).

$$\frac{1}{\eta} = \frac{1}{\eta_{\max}} + \frac{m_e}{\mu} + \frac{\xi_p}{\eta \xi_p^{\max}} \tag{7}$$

where $m_e = \mu\epsilon_m/\eta$ is the rate of consumption of energy for maintenance per unit of energy in the biomass per hour. In animal nutrition studies the maintenance energy requirement has been found to depend on the 0.75 power of the mass of the animal (Erickson and Patel, 1982; Brody, 1945).

Several variables may be measured in microbial growth processes. Biomass production, product formation, specific growth rate, and substrate consumption measurements are required for direct evaluation of the quantities η, μ, and ξ_p in Eq. 7. If oxygen uptake measurements are also made, the equation (Erickson, 1979).

$$\frac{\epsilon}{\eta} = \frac{1 - \eta_{\max}}{\eta_{\max}} + \frac{m_e}{\mu} + \frac{\xi_p}{\eta}\frac{(1 - \xi_p^{\max})}{\xi_p^{\max}} \tag{8}$$

may be used for parameter estimation where oxygen uptake, biomass production, specific growth rate, and product formation measurements are required for the direct calculation of ϵ/η, μ, and ξ_p/η. When carbon dioxide production measurements are also made, the equation (Erickson, 1979)

$$\frac{d}{\eta} = \frac{1 - \frac{\gamma_s}{\gamma_b}\eta_{\max}}{\eta_{\max}} + \frac{m_e}{\mu} + \frac{\xi_p}{\eta}\frac{\left(1 - \frac{\gamma_s}{\gamma_p}\xi_p^{\max}\right)}{\xi_p^{\max}} \tag{9}$$

may be used for parameter estimation where carbon dioxide production, biomass production, specific growth rate, and product formation are required for direct evaluation of d/η, μ, and ξ_p/η. In Eq. 9, d is the fraction of substrate carbon which is converted to CO_2; that is, the microbial growth process may be considered to follow a chemical balance equation of the form

$$CH_mO_l + aNH_3 + bO_2 = y_cCH_pO_nN_q + zCH_rO_sN_t + cH_2O + dCO_2 \tag{10}$$

where a, b, y_c, z, c and d are stoichiometric coefficients, CH_mO_l denotes the organic substrate, $CH_pO_nN_q$ refers to the elemental composition of the biomass, and $CH_rO_sN_t$ denotes the extracellular products. In Eq. 9, γ_s, γ_b, and γ_p are the reductance degrees of the substrate, biomass, and products, respectively as defined by Minkevich and Eroshin (1973).

The consistency of the data may be examined using an available electron balance which is given in Eq. 5 and a carbon balance based on Eq. 10

$$y_c + z + d = 1.0 \tag{11}$$

where y_c is fraction of substrate carbon incorporated into biomass, z is the fraction utilized in product formation and d is the fraction evolved as carbon dioxide.

In animal nutrition studies, the concept of specific growth rate is not widely used and

$$\epsilon_m = \frac{M_e M^{0.75}}{Q_T} \tag{12}$$

where M_e is the maintenance coefficient, M is the mass of the animal and Q_T is the energy in the feed which is metabolized by the animal per day. The term m_e/μ in Eqs. 7, 8, and 9 becomes

$$\frac{\epsilon_m}{\eta} = \frac{M_e M^{0.75}}{\eta Q_T} \tag{13}$$

where ηQ_T is the rate of gain of the animal in energy units. Some results of using these equations for parameter estimation with animal nutrition data are reported elsewhere by Erickson and Patel (1981). For microbial growth without any product formation, the available electron balance and the carbon balance of Eqs. 5 and 11 respectively, reduce to

$$\eta + \epsilon = 1 \tag{14}$$

$$y_c + d = 1 \tag{15}$$

Note that

$$y_c = \frac{\gamma_s}{\gamma_b}\eta \qquad \text{and } z = \frac{\gamma_p}{\gamma_b}\xi_p$$

Also for this situation, Pirt's (1965, 1975) model can be written in the forms (Ferrer and Erickson, 1979).

$$\frac{\mu}{\eta} = \frac{\mu}{\eta_{\max}} + m_e \tag{16}$$

$$Q_s = \frac{\mu}{Y_s\max} + m_s \tag{17}$$

$$Q_{O_2} = \frac{\mu}{Y_0\max} + m_o \tag{18}$$

$$Q_{CO_2} = \frac{\mu}{Y_D^{\max}} + m_D \tag{19}$$

where μ, η, η_{max} and m_e are as previously defined while Q_s, Q_{O_2} and Q_{CO_2} are respectively the specific rates of organic substrate consumption, oxygen consumption and carbon dioxide evaluation, Y_s^{max}, Y_o^{max} and Y_D^{max} are true growth yields based on substrate, oxygen and carbon dioxide respectively and m_s, m_o and m_D are maintenance coefficients based on substrate, oxygen and carbon dioxide respectively.

Recently, Roels (1983) has presented the development of these models from fundamental principles. The validity of the models and their individual application in parameter estimation has been demonstrated (Pirt, 1975; Roels, 1983; Erickson, 1979; Erickson, Minkevich and Eroshin, 1979; Oner, Erickson and Yang, 1983; Erickson and Hess, 1981; Solomon, Erickson and Yang, 1983).

When the growth models presented above are reparameterized, models relating the responses with the common parameters can be put into a type of multivariate linear model (model with several response variables) which is a special case of the general growth curve model introduced by Putthoff and Roy (1964). Using Rao's (1965, 1966, 1967) ideas of covariance adjustment and following Grizzle and Allen (1969), this special type of growth curve model can be reduced to a unit variate linear model (model with one response variable), with covariates. Therefore, standard multiple regression and residual analysis programs can be used to estimate parameters and to test hypotheses concerning the parameters.

COVARIANCE ADJUSTMENT APPROACH

Suppose that, in an experimental program, N sets of reaction conditions are run, and at each set of conditions p responses $(x_1,x_2,\ldots x_p)$ are recorded. Suppose furthermore that we can write a mathematical model for the ith response at the uth set of reaction conditions

$$x_{iu} = \xi_1 a_{1u} + \xi_2 a_{2u} + \ldots + \xi_r a_{ru} + \epsilon_{iu} \qquad (20)$$

$$i = 1,\ldots p \quad u = 1,\ldots,N;$$

where ϵ_{iu} is the error in the ith response for the uth run, $\xi = (\xi_1,\ldots,\xi_r)'$ are the r unknown parameters and a_{ju} $(j = 1,\ldots,r)$ are the values of the input variables defining the reaction conditions for the uth run.

We shall assume that the error vectors $\epsilon_u = (\epsilon_{1u},\epsilon_{2u},\ldots,\epsilon_{pu})'$ $(u = 1,\ldots N)$ are independently identically distributed according to multivariate normal distribution with zero mean vector and unknown covariance matrix Σ.

In matrix notation, we can express Eq. 20 as

$$X = \mathbf{1}\xi' A + E, \qquad (21)$$

where

$$X = \begin{bmatrix} x_{11} & x_{12}\ldots & x_{1N} \\ x_{21} & x_{22}\ldots & x_{2N} \\ \vdots & \vdots & \vdots \\ x_{p1} & x_{p2}\ldots & x_{pN} \end{bmatrix}$$

$$\mathbf{1} = (1,1,\ldots,1)'_{px1},$$

$$\xi = (\xi_1,\xi_2\ldots,\xi_r)'_{rx1},$$

$$A = \begin{bmatrix} a_{11} & a_{12}\ldots & a_{1N} \\ a_{21} & a_{22}\ldots & a_{2N} \\ \vdots & \vdots\ldots & \vdots \\ a_{r1} & a_{r2}\ldots & a_{rN} \end{bmatrix}$$

$$E = \begin{bmatrix} \epsilon_{11} & \epsilon_{12}\ldots & \epsilon_{1N} \\ \epsilon_{21} & \epsilon_{22}\ldots & \epsilon_{2N} \\ \vdots & \vdots\ldots & \vdots \\ \epsilon_{p1} & \epsilon_{p2}\ldots & \epsilon_{pN} \end{bmatrix}$$

Hence Eq. 21 is a special case of the general growth curve model considered by Putthoff and Roy (1964) and applied by Yang et al. (1982). Note that for each set of reaction conditions, the expected values of the responses, $x_{1u},x_{2u},\ldots,x_{pu}$, are equal for this model.

The model described by Eq. 20 or 21 may be reduced to a standard univariate linear model

$$\bar{y}_{.u} = \xi_1 a_{1u} + \xi_2 a_{2u} + \ldots + \xi_r a_{ru} + \bar{\epsilon}_{.u}$$

by letting

$$\bar{y}_{.u} = \sum_{i=1}^{p} x_{iu}/p$$

be the arithmetic average of the uth set of responses. However, to efficiently use all the information contained in the responses $x_{1u},x_{2u},\ldots,x_{pu}$, appropriately selected covariates, Z_i, are introduced. Let $Y = X'\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}$ be vector of dimension N and $Z = (Z_1,Z_2,\ldots,Z_{p-1}) = X'B$ be an $Nx(p-1)$ matrix where B is a $p\times(p-1)$ matrix of rank $(p-1)$ and $B'\mathbf{1} = \mathbf{0}$. Then, $E(Y) = A'\xi$ and $E(Z) = 0$. The expected value of each of the covariates, Z_i is zero because the covariates are selected such that their values are zero when $x_{1u} = x_{2u} = \ldots = x_{pu}$. This is the expected result for the model. Therefore the conditional expectation of Y given Z is $E(Y|Z) = A'\xi + Z\alpha$, where α is a vector of $(p-1)$ unknown regression coefficients. The conditional variance is

$$\sigma^2 = (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\Sigma\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1} - (\mathbf{1}'\mathbf{1})^{-1}\mathbf{1}'\Sigma B(B'\Sigma B)^{-1}B'\Sigma\mathbf{1}(\mathbf{1}'\mathbf{1})^{-1}$$
$$= (\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}$$

Now it is clear that the conditional linear model of Y conditional on Z is

$$Y = A'\xi + Z\alpha + e \qquad (22)$$

where the elements of the residual vector e are independently and identically distributed according to a normal distribution with zero mean and variance equal to σ^2. Equation (22) which may also be written in the form

$$\bar{y}_{.u} = \sum_{j=1}^{r} \xi_j a_{ju} + \sum_{i=1}^{p-1} \alpha_i z_{iu} + e_u$$

may be used with standard multiple regression methods to estimate the parameters ξ_j and α_i. The estimated values of α_i from the regression analysis determine the weighting associated with each covariate, Z_i.

Since $E(Z) = 0$, the marginal density of Y is independent of ξ. Therefore the maximum likelihood estimate of ξ under the conditional model (Eq. 22) which is also the least squares estimate based on Eq. 22 is the maximum likelihood estimate of ξ under the unconditional model (Eq. 21). Moreover, since the distribution of the pivotal quantities for making inferences on ξ under the conditional model (Eq. 22) are independent of Z, their conditional and unconditional distributions are the same. Hence the statistical inferences on ξ based upon the standard linear model theory under the conditional model (Eq. 22) are also valid for the unconditional model (Eq. 21).

The maximum likelihood estimate of ξ can be written as

$$\hat{\xi}' = (\mathbf{1}'\hat{\Sigma}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{\Sigma}^{-1}XA'(AA')^{-1} \qquad (23)$$

where the sample covariance matrix, $\hat{\Sigma} = X[I - A'(AA')^{-1}A]X'/(N-r)$, is a pXp matrix which is an unbiased estimate of Σ. Note that the ith row of $XA'(AA')^{-1}$ is the least squares estimate of ξ' based on the linear model for the ith response

$$X_{iu} = \xi_1 a_{1u} + \xi_2 a_{2u} + \ldots + \xi_r a_{ru} + \epsilon_{iu}(u = 1,\ldots,N).$$

Hence if we replace $\hat{\Sigma}$ by Σ, then any linear function $\lambda'\hat{\xi}$ is the minimum variance unbiased estimate of $\lambda'\xi$ when Σ is known. It can also be shown that (Grizzle and Allen 1969, Eq. 8)

$$\mathrm{Var}(\hat{\xi}) = (AA')^{-1}(\mathbf{1}'\Sigma^{-1}\mathbf{1})^{-1}(N-r-1)/(N-r-p)$$

An unbiased estimator for $\mathrm{Var}(\hat{\xi})$ is

$$\hat{\mathrm{Var}}(\hat{\xi}) = (AA')^{-1}\hat{\sigma}^2(N-r-1)/(N-r-p),$$

where $\hat{\sigma}^2$ is the mean square error for fitting the conditional model (22).

From Eq. 23, we see that ξ is essentially a weighted least squares estimate weighted inversely by the sample covariance matrix $\hat{\Sigma}$. Rao (1966, 1967) suggested that including only appropriately selected subsets of $Z_1, \ldots, Z_{p-1}$ as covariates in model (Eq. 22) may in some cases yield "better" estimates than including all the Z_i's. The least squares estimate of ξ based on model (Eq. 22) including only a subset of $Z_1, \ldots, Z_{p-1}$ is a maximum likelihood estimate based on Y and the selected subjset of the Z_i's but it is no longer the maximum likelihood estimate based on X. An unbiased estimate for the variance of the resulting estimate is given by

$$\hat{\text{Var}}(\hat{\xi}) = (AA')^{-1}\hat{\sigma}^2(N - r - 1)/(N - r - q - 1) \qquad (24)$$

where q is the total number of Z_i included in model (Eq. 22) and $\hat{\sigma}^2$ is the mean square error for fitting the reduced model (Eq. 22).

The measure of "goodness" of the selection may be based on the generalized variance of $\hat{\xi}$, det [$\hat{\text{Var}}(\hat{\xi})$]. From Eq. 24, we see that $\hat{\text{Var}}(\hat{\xi})$ depends on the selected Z_i's only through $\hat{\sigma}^2$ and q. Hence a simple measure of "goodness" of the selection is $\hat{\sigma}^2/(N - r - q - 1)$. The selection can be determined by examining the correlations between Y and $Z_i (i = 1, \ldots, p - 1)$. One can also use the dominant right hand characteristic vectors (characteristic vectors associated with "large" characteristic values) of $[I - 11'(1'1)^{-1}]\hat{\Sigma}$ as the column of B for generating the covariates Z for model (Eq. 22). These ideas will be illustrated in the numerical examples considered later.

Consider Pirt's (1965, 1975) model discussed earlier; when nitrogen, substrate, oxygen, carbon dioxide and biomass measurements are available at least four different types of true growth yields and maintenance coefficients can be independently estimated. However using the concept of available electrons and the relationships (Erickson et al., 1979; Ferrer and Erickson, 1979) between $Y_s^{\max}$, $Y_o^{\max}$, $Y_D^{\max}$, and $\eta_{\max}$ and the relationship between m_s, m_o, m_D and m_e, the true growth yields and maintenance parameters are unified and Eqs. 16, 17, 18 and 19 can be reparameterized to the form

$$\frac{\mu}{\eta_{N_2}} = \frac{\mu}{\eta_{\max}} + m_e + \epsilon_1 \qquad (25)$$

$$\frac{\mu}{\eta} = \frac{\sigma_s\gamma_s}{\sigma_b\gamma_b} Q_s = \frac{\mu}{\eta_{\max}} + m_e + \epsilon_2 \qquad (26)$$

$$\frac{\mu}{\eta}(\eta + \epsilon) = \frac{\mu}{\eta_{\max}} + m_e + \epsilon_3 \qquad (27)$$

$$\frac{\mu}{\eta}(y_c + d) = \frac{\mu}{\eta_{\max}} + m_e + \epsilon_4 \qquad (28)$$

where σ_s and σ_b are the fractions of carbon in substrate and biomass respectively, γ_s and γ_b are the reductance degrees of organic substrate and biomass respectively, and ϵ_1, ϵ_2, ϵ_3, and ϵ_4 are the random errors. The dependent variables μ/η_{N_2} and μ/η from Eqs. 25 and 26 respectively are obtained from nitrogen, biomass and substrate, and substrate and biomass measurements. The dependent variables in Eqs. 27 and 28 involve the available electron balance $(\eta + \epsilon)$ and the carbon balance $(y_c + d)$ respectively. For consistent data the constraints $\eta + \epsilon = 1$ and $y_c + d = 1$ are satisfied and Eqs. 26, 27 and 28 are essentially identical. However since the available electron and carbon balances are rarely satisfied, due to measurement errors, different estimations of $\eta_{\max}$ and m_e can be obtained from the above equations. If the data is statistically adjusted such that it satisfies the available electron balance and carbon balance constraints then only one estimate to the parameters can be obtained from the three equations. However our approach has been not to adjust the data to satisfy the constraints but instead to use the dependent variables which are related by the available electron balance and carbon balance as correlated multiresponse observations.

The columns of the data matrix X of model (Eq. 21) are the N observed responses

$$\left[\frac{\mu}{\eta_{N_2}}, \quad \frac{\mu}{\eta}, \quad \mu\frac{(\eta + \epsilon)}{\eta}, \quad \mu\left(\frac{y_c + d}{\eta}\right)\right]'$$

The A matrix is

$$A = \begin{bmatrix} 1 & 1 \ldots 1 \\ \mu_1 & \mu_2 \ldots \mu_N \end{bmatrix}$$

where $\mu_i (i = 1, \ldots, N)$ are the observed specific growth rates. The parameter vector is

$$\xi = \begin{bmatrix} m_e \\ \dfrac{1}{\eta_{\max}} \end{bmatrix}$$

For convenience, in almost all cases, we choose B to be a matrix with columns which are coefficients of the orthogonal polynomials. In this example with four responses

$$B = \begin{bmatrix} -3 & 1 & -1 \\ -1 & -1 & 3 \\ 1 & -1 & -3 \\ 3 & 1 & 1 \end{bmatrix}$$

The covariates are defined as

$$Z = (Z_1, Z_2, Z_3) = X'B.$$

$$Y = X' \begin{bmatrix} 1/4 \\ 1/4 \\ 1/4 \\ 1/4 \end{bmatrix}$$

The corresponding conditional model is

$$Y = \begin{bmatrix} 1 & \mu_1 \\ 1 & \mu_2 \\ \vdots & \vdots \\ 1 & \mu_N \end{bmatrix} \begin{pmatrix} m_e \\ \dfrac{1}{\eta_{\max}} \end{pmatrix} + \sum_{i=1}^{3} \alpha_i Z_i + e. \qquad (29)$$

For the extracellular product formation case, Eqs. 7, 8 and 9 may be rearranged and put into linearized form as follows:

$$1/\eta = 1/\eta_{\max} + \frac{m_e}{\mu} + \frac{1}{\xi_p^{\max}}\left(\frac{\xi_p}{\eta}\right) + \epsilon_1 \qquad (30)$$

$$\frac{\epsilon + \eta + \xi_p}{\eta} = 1/\eta_{\max} + \frac{m_e}{\mu} + \frac{1}{\xi_p^{\max}}\left(\frac{\xi_p}{\eta}\right) + \epsilon_2 \qquad (31)$$

$$\frac{y_c + z + d}{\eta} = 1/\eta_{\max} + \frac{m_e}{\mu} + \frac{1}{\xi_p^{\max}}\left(\frac{\xi_p}{\eta}\right) + \epsilon_3 \qquad (32)$$

Again the dependent observations of Eqs. 31 and 32 involve the available electron balance and the carbon balance. Also when the constraints are satisfied $\eta + \epsilon + \xi_p = 1$ and $y_c + d + z = 1$ and Eqs. 30, 31 and 32 are identical. When no data adjustment is made, this can be treated as a multiresponse observation and the covariance adjustment technique can be used for estimation of the parameters.

NONLINEAR ESTIMATION PROCEDURES

Equations 7, 8 and 9 will be referred to as Form I and used with nonlinear parameter estimation methods to obtain estimates of $\eta_{\max}$, m_e, and $\xi_p^{\max}$. Multiplication of each term in Eqs. 7, 8 and 9, by μ gives Form II:

$$\frac{\mu}{\eta} = \frac{\mu}{\eta_{\max}} + m_e + \frac{\xi_p\mu}{\eta}\frac{1}{\xi_p^{\max}} \qquad (33)$$

$$\frac{\epsilon\mu}{\eta} = \frac{1 - \eta_{\max}}{\eta_{\max}} \cdot \mu + m_e + \frac{\xi_p\mu}{\eta}\left(\frac{1 - \xi_p^{\max}}{\xi_p^{\max}}\right) \qquad (34)$$

$$\frac{d\mu}{\eta} = \frac{\left(1 - \dfrac{\gamma_s}{\gamma_b}\eta_{\max}\right)}{\eta_{\max}} \cdot \mu + m_e + \frac{\xi_p\mu}{\eta}\left(\frac{1 - \dfrac{\gamma_s}{\gamma_p}\xi_p^{\max}}{\xi_p^{\max}}\right) \qquad (35)$$

Form I and Form II correspond to the two different forms considered previously in yield studies without products (Pirt, 1965; Hempfling and Mainzer, 1975). Because of measurement errors, parameter estimates differ for each of these six equations. Equations 7 and 33 give different parameter estimates because the forms of the two equations differ while Eqs. 7, 8 and 9 give different parameter estimates because different experimental data is used. Biomass, product, organic substrate, and specific growth rate measurements are used with Eq. 7 or 33. Oxygen, biomass, product and specific growth rate measurements are used with Eq. 8 or 34. Carbon dioxide, biomass, product, and specific growth rate measurements are used with Eq. 9 or 35.

Several different methods of nonlinear statistical treatment of the data may be employed to obtain point and interval estimates of m_e, $\eta_{\max}$, and $\xi_p^{\max}$. Two different methods are employed in this work and the results are compared (Oner et al., 1983).

Nonlinear Method I

Equations 7–9 and 33–35 can be rearranged in the following form:

$$Y_{1i} = \frac{1}{\eta_{\max}} + \frac{m_e}{X_{1i}} + X_{2i}\frac{1}{\xi_p^{\max}} + e_{1i} \tag{36}$$

$$Y_{2i} = \frac{1-\eta_{\max}}{\eta_{\max}} + \frac{m_e}{X_{1i}} + X_{2i}\frac{1-\xi_p^{\max}}{\xi_p^{\max}} + e_{2i} \tag{37}$$

$$Y_{3i} = \frac{1-\frac{\gamma_s}{\gamma_b}\eta_{\max}}{\eta_{\max}} + \frac{m_e}{X_{1i}} + X_{2i}\frac{1-\frac{\gamma_s}{\gamma_p}\xi_p^{\max}}{\xi_p^{\max}} + e_{3i} \tag{38}$$

$$Y_{4i} = \frac{1}{\eta_{\max}}X_{1i} + m_e + X'_{2i}\frac{1}{\xi_p^{\max}} + e_{4i} \tag{39}$$

$$Y_{5i} = \frac{1-\eta_{\max}}{\eta_{\max}}X_{1i} + m_e + X'_{2i}\frac{1-\xi_p^{\max}}{\xi_p^{\max}} + e_{5i} \tag{40}$$

$$Y_{6i} = \frac{1-\frac{\gamma_s}{\gamma_b}\eta_{\max}}{\eta_{\max}}X_{1i} + m_e + X'_{2i}\frac{1-\frac{\gamma_s}{\gamma_p}\xi_p^{\max}}{\xi_p^{\max}} + e_{6i} \tag{41}$$

where;

$$Y_{1i} = \frac{1}{\eta_i}, \quad Y_{2i} = \frac{\epsilon_i}{\eta_i}, \quad Y_{3i} = \frac{d_i}{\eta_i}$$

$$Y_{4i} = \frac{\mu_i}{\eta_i}, \quad Y_{5i} = \frac{\epsilon_i\mu_i}{\eta_i}, \quad Y_{6i} = \frac{d_i\mu_i}{\eta_i}$$

$$X_{1i} = \mu_i, \quad X_{2i} = \frac{\xi_{pi}}{\eta_i}, \quad X'_{2i} = \frac{\xi_{pi}\mu_i}{\eta_i}$$

$i = 1,2,\ldots,N$ where N is the number of data points. The values of reductance degree γ_s, γ_b and γ_p, are constants for specified substrate, biomass and extracellular products, respectively (Erickson et al., 1978, 1979). The last term in Eqs. 36–41, e_{ji}, is an error term. It can be defined as:

$$e_{ji} = Y_{ji,\text{obs}} - Y_{ji,\text{calculated}} \tag{42}$$

One can find the estimates of parameters m_e, $\eta_{\max}$ and $\xi_p^{\max}$ by using a nonlinear least square technique which is based on minimizing the quantity (Draper and Smith, 1966).

$$\sum_{i=1}^{N}[e_{ji}]^2 \text{ for } j = 1,2,\ldots,6 \tag{43}$$

Also, it is possible to obtain combined estimates of parameters from M of Eqs. 7–9 or 33–35 by minimizing (Hunter, 1967; Ball and Groenweghe, 1966).

$$\sum_{j=1}^{M}\sum_{i=1}^{N}(e_{ji})^2 \tag{44}$$

for M equations each with N data points.

For the case of correlated Y_{ji} and $Y_{j'i}(j \neq j')$, Box and Draper (1965) have derived a determinant criterion which, when minimized with respect to the parameters, m_e, $\eta_{\max}$, and $\xi_p^{\max}$ gives estimates of these unknown parameters in the multiresponse models (Eqs. 26–38 or 39–41); that is, the determinant of the following matrix is minimized.

$$\begin{bmatrix} \sum_{i=1}^{N}(e_{1i})^2 & \sum_{i=1}^{N}(e_{1i})(e_{2i}) & \cdots & \sum_{i=1}^{N}(e_{1i})(e_{Mi}) \\ \sum_{i=1}^{N}(e_{2i})(e_{1i}) & \sum_{i=1}^{N}(e_{2i})^2 & \cdots & \sum_{i=1}^{N}(e_{2i})(e_{Mi}) \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N}(e_{Mi})(e_{1i}) & \sum_{i=1}^{N}(e_{Mi})(e_{2i}) & \cdots & \sum_{i=1}^{N}(e_{Mi})^2 \end{bmatrix} \tag{45}$$

In this work, the Hooke and Jeeves Pattern Search Technique (1961) was used to estimate parameters by minimizing the above determinant.

For 95% confidence intervals of parameters, the Smooth Bootstrap Method was used (Efron, 1979, 1981).

Basically, N sets of data are generated based on the assumed model and the observed data using Monte Carlo simulation techniques. From the N sets of generated data, N estimates of the parameters are obtained. These estimates are arranged in ascending order as for example,

$$\hat{\eta}_{\max}^1 \leq \hat{\eta}_{\max}^2 \leq \cdots \leq \hat{\eta}_{\max}^N \tag{46}$$

Then a 95% confidence interval for $\eta_{\max}$ is $[\eta_{\max}^r, \eta_{\max}^{N-r}]$ where r is the integral part of $0.025\,N$. Of course, the larger the value of N the more accurate the estimated interval is.

Nonlinear Method II

As a first step, nonlinear regression was applied to each of the Eqs. 36–38 or 39–41 to obtain least squares estimates of errors $(\hat{e}_{ji})$ by using SAS (1979).

These estimates of errors were used to estimate the variance covariance matrix (Σ) as follows:

$$\Sigma = \frac{1}{N-P}\begin{bmatrix} \sum_{i=1}^{N}(\hat{e}_{1i})^2 & \sum_{i=1}^{N}(\hat{e}_{1i})(\hat{e}_{2i}) & \cdots & \sum_{i=1}^{N}(\hat{e}_{1i})(\hat{e}_{Mi}) \\ \sum_{i=1}^{N}(\hat{e}_{2i})(\hat{e}_{1i}) & \sum_{i=1}^{N}(\hat{e}_{2i})^2 & \cdots & \sum_{i=1}^{N}(\hat{e}_{2i})(\hat{e}_{Mi}) \\ \vdots & \vdots & & \vdots \\ \sum_{i=1}^{N}(\hat{e}_{Mi})(\hat{e}_{1i}) & \sum_{i=1}^{N}(\hat{e}_{Mi})(\hat{e}_{2i}) & \cdots & \sum_{i=1}^{N}(\hat{e}_{Mi})^2 \end{bmatrix} \tag{47}$$

where P is the number of parameters to be estimated.

Combined point estimates were obtained by minimizing the following quantity:

$$\sum_{i=1}^{N}(e_{1i},e_{2i},\ldots,e_{Mi})\,\Sigma^{-1}\begin{Bmatrix} e_{1i} \\ e_{2i} \\ \vdots \\ e_{Mi} \end{Bmatrix} \tag{48}$$

where e_{ji} is given by Eq. 42; that is

$$e_{ji} = y_{ji,\text{obs}} - F_j(X_{1i},X_{2i}\ (\text{or } X'_{2i}),m_e,\eta_{\max},\xi_p^{\max}) \tag{42}$$

Again, the Hooke and Jeeves Pattern Search Technique (1961) was used.

This method is principally weighted least squares based on the estimated covariance matrix. The 95% confidence intervals were estimated with the same procedure of Method I.

RESULTS AND DISCUSSION

The parameters m_e, $\eta_{\max}$, and $\xi_p^{\max}$ are estimated using the methods described above. Only positive values of these parameters are physically meaningful. Values of the maintenance coefficient are often small, and it is common for the 95% confidence interval to include zero. Values of $\eta_{\max}$ and $\xi_p^{\max}$ should be less than one; the

TABLE 1. ESTIMATES OF TRUE BIOMASS ENERGETIC YIELD AND MAINTENANCE PARAMETER FROM BIOMASS, DILUTION RATE, OXYGEN AND SUBSTRATE DATA FOR THE GROWTH OF *PARACOCCUS DENITRIFICANS*

References	Culture Medium	ρ_{yZ_1}	N	Z_1 Included As a Covariate?	$\hat{\sigma}^2$ / N-r-q-1	η_{max} Point	η_{max} Interval	m_e, h^{-1} Point	m_e, h^{-1} Interval
Meijer et al. (1977)	Gluconate limited	−0.740	17	Yes	$9.026 * 10^{-6}$	0.634	[0.601, 0.672]	0.017	[0.002, 0.031]
"	"	"	"	No	$1.174 * 10^{-5}$	0.605	[0.579, 0.633]	0.016	[−0.002, 0.033]
Meijer et al. (1977)	Gluconate with sulfate limitation	−0.589	11	Yes	$1.245 * 10^{-5}$	0.547	[0.509, 0.592]	0.029	[0.012, 0.047]
"	"	"	"	No	$2.175 * 10^{-5}$	0.520	[0.480, 0.567]	0.036	[0.013, 0.059]
Meijer et al. (1977)	Succinate limited	−0.871	14	Yes	$1.927 * 10^{-5}$	0.585	[0.531, 0.652]	0.036	[0.017, 0.056]
"	"	"	"	No	$3.517 * 10^{-5}$	0.513	[0.479, 0.552]	0.037	[0.009, 0.064]
Meijer et al. (1977)	Succinate with sulfate limitation	−0.436	13	Yes	$1.131 * 10^{-5}$	0.464	[0.440, 0.490]	0.049	[0.031, 0.068]
"	"	"	"	No	$1.970 * 10^{-5}$	0.450	[0.442, 0.482]	0.066	[0.046, 0.087]
Van Verseveld and Stouthamer (1976)	Malate limited	−0.208	10	Yes	$3.694 * 10^{-4}$	0.581	[0.545, 0.621]	0.049	[0.024, 0.074]
"	"	"	"	No	$2.822 * 10^{-5}$	0.579	[0.548, 0.615]	0.050	[0.027, 0.072]
Van Verseveld & Stouthamer	Mannitol limited	− 739	13	Yes	$2.653 * 10^{-5}$	0.611	[0.563, 0.669]	0.016	[−0.006, 0.039]
"	"	"	"	No	$4.046 * 10^{-5}$	0.567	[0.527, 0.613]	0.010	[−0.018, 0.038]

$Z_1 = [-1\ 1]\ X$

biochemical maximum for η_{max} appears to be about 0.88 (Erickson, 1980).

The data in Table 1 represents processes with two responses. It contains the estimates of the true biomass energetic yield, η_{max} and maintenance, m_e obtained using the direct measurement of biomass and substrate only in Eq. 26 and the values of the available electron balance in Eq. 27 for the growth of *Paracoccus dentrificans* on several culture media. The estimates were obtained using the covariance adjustment technique. It was observed that point and interval estimates obtained when the covariates were included were identical to maximum likelihood estimates reported in an earlier paper (Solomon et al., 1982). As discussed in the theory section the magnitudes of the correlation coefficient (ρ_{yZ_1}), $\hat{\sigma}^2/(N - r - q - 1)$ and the interval estimates are to be examined and compared. When the value of the correlation coefficient is high, inclusion of the covariate may be advantageous. Examination of the results in Table 1 shows that when the value of the correlation coefficient is large in magnitude, the smallest values of $\hat{\sigma}^2/(N - r - q - 1)$ are associated with the results where the covariate Z_1, is included. Also the shortest confidence interval for m_e is always associated with the result which gave the smallest value of $\hat{\sigma}^2/(N - r - q - 1)$. For the malate limited culture, best results are obtained when the covariate Z_1, is not included (that is the MLE is not the best estimator); this is what one would expect based on both the small value of the correlation coefficient and the fact that the smaller value of $\hat{\sigma}^2/(N - r - q - 1)$ occurs when the covariate is not included.

The regression results are shown graphically in Figures 1 and 2 for the malate limited culture. The regression line which is shown represents both the MLE result and the result without any covariates (the two curves are so close to each other that only one line is shown). In Figure 1, the circles correspond to the lefthand side of Eq. 26 which is obtained from biomass and substrate measurements. Biomass and oxygen measurements are used to obtain the values of $(\eta + \epsilon)\mu/\eta$ for Eq. 27 which are shown as squares in Figure 1. The regression line is obtained by using the estimated values of the parameters m_e and η_{max} to calculate

$$\hat{y}_{.i} = m_e + \frac{\mu_i}{\eta_{max}} \tag{49}$$

The residual plot corresponding to Figure 1 is shown in Figure 2 where $\bar{y}_{.i} - \hat{y}_{.i}$ is the residual. For the MLE case in which a covariate is included, the term involving the covariate, $\alpha_1 Z_{1i}$, is added to the righthand side of Eq. 49 in calculating the predicted value. The distribution of points on the residual plot supports the validity of the model. The small differences between the two regression

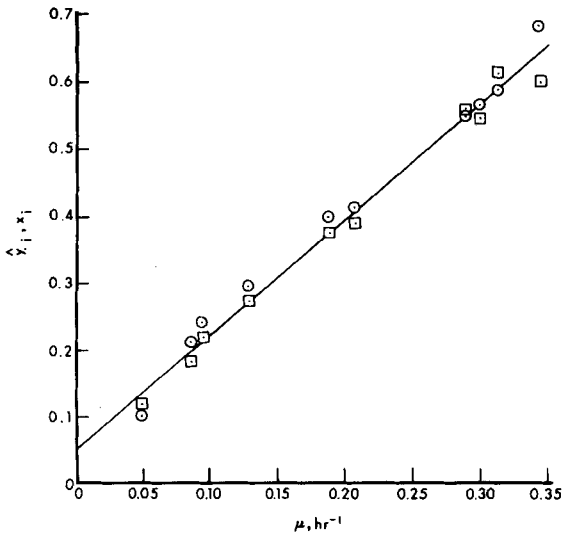

Figure 1. Regression lines and data for growth of *Paracoccus denitrificans* in malate limited continuous culture. Regression lines with covariates and without covariates are represented by the same line. Circles represent data based on biomass and substrate measurements while squares represent data based on biomass and oxygen measurements.

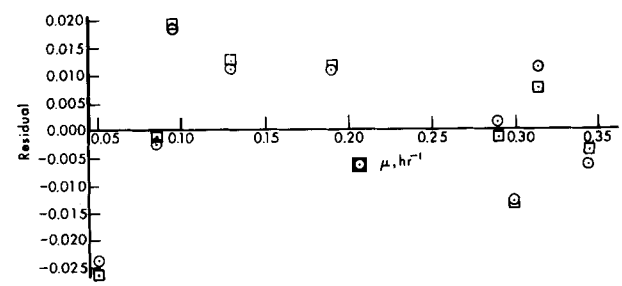

Figure 2. Residual plot of $\bar{y}_{.i} - \hat{y}_{.i}$ for *Paracoccus denitrificans* in malate limited culture for data in Figure 1. Circles represent residuals with covariate adjustment while squares are for the case of no covariates.

TABLE 2. ESTIMATES OF TRUE BIOMASS ENERGETIC YIELDS AND MAINTENANCE PARAMETERS FROM BIOMASS, SPECIFIC GROWTH RATE, O_2, CO_2, GLUCOSE AND NITROGEN DATA FOR THE GROWTH OF *CANDIDA UTILIS* ON FILTERED AND UNFILTERED GLUCOSE FROM CORN DUST

Medium	Covariates Included	$\hat{\sigma}^2$ / N-r-q-1	η_{max} Point	η_{max} Interval	m_e, h^{-1} Point	m_e, h^{-1} Interval
a	—	$9.375 * 10^{-5}$	0.616	[0.552, 0.695]	0.045	[−0.033, 0.123]
b	—	"	0.557	[0.509, 0.616]	0.004	[−0.080, 0.088]
a	Z_1	$9.848 * 10^{-5}$	0.616	[0.552, 0.696]	0.044	[−0.034, 0.123]
b	Z_1	"	0.573	[0.513, 0.649]	0.017	[−0.072, 0.106]
a	Z_2	$8.091 * 10^{-5}$	0.612	[0.554, 0.683]	0.034	[−0.038, 0.106]
b	Z_2	"	0.668	[0.559, 0.829]	0.098	[−0.0108, 0.206]
a	Z_3	$1.009 * 10^{-4}$	0.622	[0.553, 0.712]	0.046	[−0.034, 0.126]
b	Z_3	"	0.571	[0.500, 0.667]	0.008	[−0.079, 0.095]
a	Z_1,Z_2	$8.414 * 10^{-5}$	0.610	[0.553, 0.681]	0.030	[−0.043, 0.102]
b	Z_1,Z_2	"	0.693	[0.571, 0.882]	0.119	[−0.003, 0.235]
a	Z_1,Z_3	$1.066 * 10^{-4}$	0.606	[0.532, 0.704]	0.043	[−0.038, 0.123]
b	Z_1,Z_3	"	0.563	[0.491, 0.659]	0.019	[−0.072, 0.111]
a	Z_2,Z_3	$8.773 * 10^{-5}$	0.605	[0.543, 0.683]	0.032	[−0.042, 0.105]
b	Z_2,Z_3	"	0.661	[0.550, 0.827]	0.102	[−0.010, 0.214]
a	Z_1,Z_2,Z_3	$9.032 * 10^{-5}$	0.625	[0.552, 0.720]	0.031	[−0.043, 0.104]
b	Z_1,Z_2,Z_3	"	0.739	[0.564, 1.07]	0.128	[0.007, 0.249]
a[1]	Z^*	$8.47 * 10^{-5}$	0.608	[0.550, 0.679]	0.030	[−0.043, 0.103]
b[1]	Z^*	$8.47 * 10^{-5}$	0.683	[0.567, 0.859]	0.115	[0.00003, 0.231]

Medium a is filtered glucose; medium b is unfiltered glucose; $\rho_{yZ_1} = 0.507$, $\rho_{yZ_2} = 0.700$, $\rho_{yZ_3} = 0.704$, $\rho_{Z_1Z_2} = 0.823$, $\rho_{Z_1Z_3} = 0.872$, $\rho_{Z_2Z_3} = 0.782$.

$$Z = [Z_1, Z_2, Z_3]' = \begin{bmatrix} -3 & -1 & 1 & 3 \\ 1 & -1 & -1 & 1 \\ -1 & 3 & -3 & 1 \end{bmatrix} X$$

[1] Analyzed using the right eigenvector of $[I - 11'(1'1)^{-1}]\hat{\Sigma}$ with dominant eigenvalues.

$$Z^{*\prime} = \begin{bmatrix} -0.3096 & -0.3420 & -0.2103 & 0.8619 \\ 0.8083 & -0.4623 & -0.3641 & 0.0180 \end{bmatrix} X$$

Data of Solomon et al. (1981). $N = 15$ for medium a; $N = 14$ for medium b.

lines are supported by the similarities of the two sets of residuals.

Contained in Table 2 are the results of covariance adjustment technique for the analysis on the data obtained for the growth of *Candida utilis* on filtered and unfiltered glucose obtained by enzymatic hydrolysis of corn dust (Solomon et al., 1981). All of the data for medium a (filtered) and medium b (unfiltered) were combined to obtain the values of the correlation coefficients and $\hat{\sigma}^2/(N - r - q - 1)$. In this analysis all nitrogen, biomass, substrate, oxygen and carbon dioxide data were used. Thus this is a situation with four responses per observation; the available electron and carbon balances were used in addition to the direct measure of biomass and substrate only and the indirect biomass estimates obtained from nitrogen and substrate measurements. The data and the consistency of the data have been reported by Solomon et al. (1981).

The results in Table 2 show that the inclusion of covariate Z_2 alone in the model is the best since the value of $\hat{\sigma}^2/(N - r - q - 1) = 8.091 * 10^{-5}$ is the lowest of all other combinations. However looking at the correlation coefficient it can be seen that all of the covariates Z_1, Z_2, and Z_3 are highly correlated with Y and are also highly correlated with each other. Based on the correlation coefficients, one would tend to assume that the best model should correspond to the results when all the covariates are included; however, because the covariates themselves are highly correlated, inclusion will not be advantageous since the information in just one or two could be as good as that obtained from all. Examination of Table 2 shows that the selection of the best estimate could become very cumbersome as the number of covariates increase; this is because the number of combinations increase very rapidly with number of responses. For example, with two responses, we have two alternatives; with three responses, there are four alternatives; while with four responses, eight different alternatives need to be explored. Instead of going through this boring process of examining various combinations, a more efficient method discussed in the theory is also tried. The three nonzero characteristic values of $[I - 11'(1'1)^{-1}]\hat{\Sigma}$ are (0.4096, 0.0777, 0.0651). The corresponding characteristic vectors are respectively (−0.3096, −0.3420, −0.2103, 0.8619),(0.8083, −0.4623, −0.3641, 0.0181), and (0.0275, 0.6475, −0.7571, 0.0821). The two characteristic vectors corresponding to the two dominant characteristic values are chosen as the two columns of B. The results obtained are very similar to those obtained including only Z_2 as covariate. In fact, the "dominant characteristic vectors" method gives shorter confidence intervals for η_{max} and m_e for growth on filtered glucose (Table 2).

Figures 3, 4, and 5 graphically illustrate some of the results reported in Table 2. Figure 3 shows the regression lines for the case where the hydrolyzed corn dust is filtered prior to using the filtrate as substrate. As shown in Figure 3 the regression line when no covariates are used (solid line) and the regression line with the covariate Z^* (dashed line) differ sufficiently that two separate lines

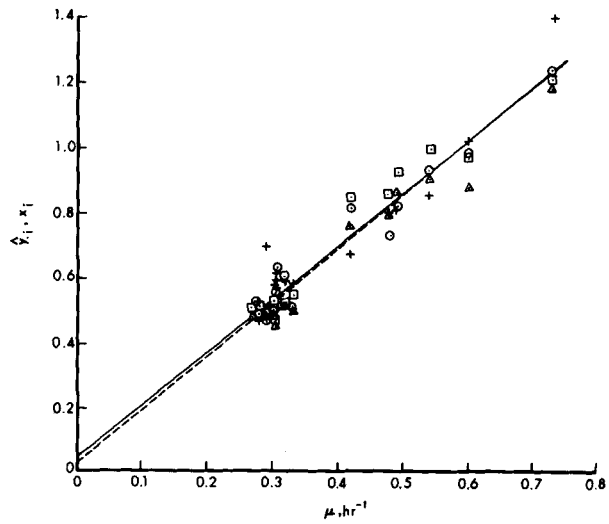

Figure 3. Regression results for Equation (29) and data for growth of *Candida utilis* on filtered glucose from corn dust. Lines correspond to Table 2 results for no covariates (solid line) and covariate Z^* (dashed line). Circles correspond to substrate and biomass data; squares correspond to oxygen and biomass data; triangles correspond to CO_2 and biomass data; plus signs correspond to nitrogen, substrate, and biomass data.

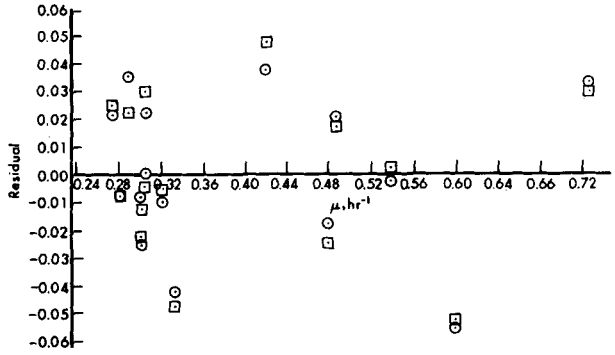

Figure 4. Residual plot of $\bar{y}_{.i} - \hat{y}_{.i}$ for *Candida utilis* for data shown in Figure 3. Circles represent residuals with covariate adjustment while squares are for the case of no covariates.

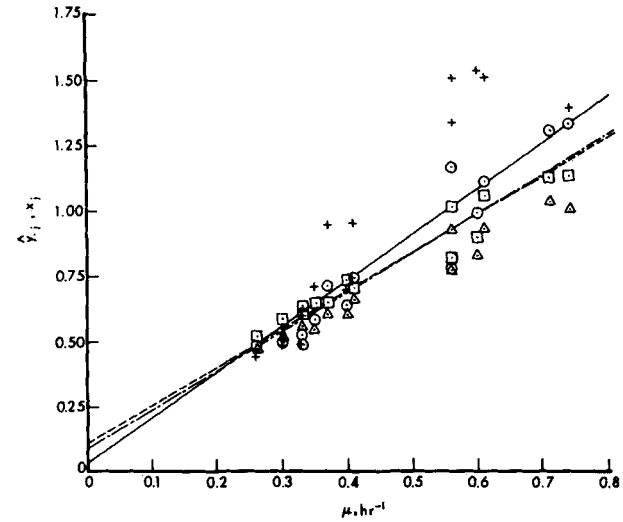

Figure 5. Regression results for Equation (29) and data for growth of *Candida utilis* on unfiltered glucose from corn dust. Lines correspond to Table 2 results with no covariates (solid line), Covariate Z* (dashed line), and Z_2 (long and short dashes). Data symbols as in Figure 3.

may be drawn. In each case the parameter estimates are used with Eq. 49 to draw the regression lines. The ordinate associated with the data is the left hand side of Eq. 25 for the plus signs, Eq. 26 for the circles, Eq. 27 for the squares and Eq. 28 for the triangles.

Figure 4 shows the residuals $(\bar{y}_{.i} - \hat{y}_{.i})$ for the results in Figure 3. For the case in which covariates are included, Eq. 29, with the estimated values of the parameters, is used to calculate $\hat{y}_{.i}$; that is, the terms involving the covariates are included in the right hand side. The points on the residual plot appear to be nearly randomly distributed. The magnitude of the residuals with and without covariates is similar in a number of cases.

Figure 5 shows the regression line and data for growth on medium b in which the residual solids are not filtered out. The solid line for the case in which no covariates are used in parameter estimation has a larger slope than the two dashed lines in which covariate adjustment was used. This is because the relatively scattered nitrogen data is weighted more heavily when no covariates are included than it is when covariate adjustment is used.

In Figures 3 and 5, it is clear that additional data at low specific growth rates would be desirable. This is difficult to obtain in batch culture, but batch culture followed by fed batch culture may be used to enhance data collection at low specific growth rates (Solomon et al., 1983).

Tables 3 and 4 compare the point and interval estimates obtained by different estimation procedures for the true biomass energetic yield and maintenance coefficient for *Candida utilis* growth on filtered glucose. In Table 3, estimates were obtained using direct biomass and substrate measurements, available electron balance and carbon balance; thus there are three responses for each observation. The point and interval estimates for both η_{max} and m_e are quite similar for the methods. In Table 4, estimates were obtained using the nitrogen measurement in addition to the data used in Table 3. The estimates obtained for η_{max} and m_e are quite similar in the two cases when the linear estimation procedures, covariance adjustment and MLE were applied. However, due to the increment in the number of responses, the size of the matrix whose determinant is to be minimized by Hooke and Jeeves search technique increases; thus, the estimates obtained, when the nonlinear procedures were used, were quite sensitive to the step sizes and starting points and also required a lot of computing time. This is explained in terms of the presence of several local minimum as the number of responses increase.

Tables 5 and 6 compare the estimates obtained using the estimation procedures for growth on unfiltered glucose. Due to the presence of unhydrolysed corn dust residues the measurement of biomass was difficult and thus the data was inconsistent (Solomon et al., 1981). The estimates of η_{max} and m_e in Table 6 using the covariance adjustment technique appear to be substantially improved when the nitrogen data is included. The values of $\eta_{max} = 0.683$ and $m_e = 0.098\ h^{-1}$ are very reasonable and the 95% confidence intervals of [0.567, 0.859] for η_{max} and [0.00, 0.231] for m_e are such that one would reasonably expect these intervals to include the actual values of the parameters. These values are considerably better than the estimates obtained previously by Solomon et al. (1981). Due to the inconsistency none of the methods could yield a very good result in Table 5. In Table 6, the nonlinear methods again had their problems due to the large number of responses at each observation.

Table 7 contains the estimate of true biomass energetic yield, η_{max}, true product yield ξ_p^{max} and the maintenance coefficients m_e, for a growth process with product formation (Erickson and Hess, 1981). In this case the microorganism *Rhizobium trifolii* produced polysaccharide in a chemostat culture. The results in Table 7 show that the point estimates of the true biomass energetic yield, η_{max}, are very consistent for nonlinear methods I and II and the covariance adjustment technique for Form I and Form II, but that the estimates for Form I are larger than those for Form II. The interval estimates for η_{max} differ considerably; nonlinear method II gives the shortest interval for Form I while nonlinear method I gives the shortest interval for Form II. These two intervals are similar. The covariance adjustment procedure gives the largest 95% confidence intervals. Note that in this analysis Eqs. 7, 8, 30, 31, 33 and 34 are used; the carbon dioxide data in the carbon balance needed for Eqs. 9, 32 and 35 is not available. Thus, there are only two responses.

The point estimates for the true product energetic yield ξ_p^{max} are very similar in Table 7 for nonlinear methods I and II and covariance adjustment technique for both Forms I and II. The interval estimates include an interval which exceeds the maximum theoretical value of 0.93 (Erickson and Hess, 1981) except for Form I, nonlinear method II. The interval of interest is from the lower limit to 0.93. The shortest interval $0.61 \leq \xi_p^{max} \leq 0.89$ from Form I, nonlinear method II appears to be a realistic interval.

The point estimates for the maintenance coefficient, m_e, are all relatively small for all the methods and both forms. All of the interval estimates include an interval where m_e is negative; however, this can be removed because m_e must be positive based on thermodynamic considerations. The interval estimates show that m_e is small for this organism.

The nonlinear estimation methods utilize the Bootstrap method which does not require any distributional assumptions with regard to the model of the data in constructing the confidence interval. Because of the simulation approach which is used to obtain the confidence intervals the interval may deviate slightly from a 95% confidence interval. The accuracy of the interval depends on the number of simulations which might require a lot of computer time. To use the covariance adjustment method, the equations need to be linearized; this linearization of the equations may affect the estimator.

When only two or three responses are observed, the nonlinear

TABLE 3. ESTIMATES OF TRUE BIOMASS ENERGETIC YIELD AND MAINTENANCE COEFFICIENTS FROM BIOMASS, O_2, CO_2 AND GLUCOSE MEASUREMENT FOR THE GROWTH OF *CANDIDA UTILIS* ON FILTERED GLUCOSE FROM CORN DUST

Method of Analysis	η_{max}		m_e, h^{-1}	
	Point	Interval	Point	Interval
Cov. Adjustment	0.615	[0.559, 0.682]	0.020	[−0.047, 0.086]
MLE	0.643	[0.574, 0.731]	0.028	[−0.038, 0.094]
Nonlin. Method I	0.657	[0.625, 0.706]	0.027	[−0.010, 0.073]
Nonlin. Method II	0.643	[0.580, 0.717]	0.026	[0.020, 0.035]

Data of Solomon et al. (1981). $N = 15$.

TABLE 4. ESTIMATES OF TRUE BIOMASS ENERGETIC YIELDS AND MAINTENANCE COEFFICIENTS FROM BIOMASS, O_2, CO_2, NITROGEN AND GLUCOSE MEASUREMENT FOR THE GROWTH OF *CANDIDA UTILIS* ON FILTERED GLUCOSE FROM CORN DUST

Method of Analysis	η_{max}		m_e, h^{-1}	
	Point	Interval	Point	Interval
Cov. Adjustment	0.608	[0.550, 0.679]	0.030	[−0.043, 0.103]
MLE	0.625	[0.553, 0.720]	0.031	[−0.047, 0.104]
Nonlin. Method I	0.578	[0.500, 0.670]	0.059	[0.041, 0.062]
Nonlin. Method II	0.589	[.554, 0.610]	0.060	[0.053, 0.067]

Data of Solomon et al. (1981). $N = 15$.

TABLE 5. ESTIMATES OF TRUE BIOMASS ENERGETIC YIELD AND MAINTENANCE COEFFICIENTS FROM BIOMASS, O_2, CO_2 AND GLUCOSE MEASUREMENT FOR THE GROWTH OF *CANDIDA UTILIS* ON UNFILTERED GLUCOSE FROM CORN DUST

Method of Analysis	η_{max}		m_e, h^{-1}	
	Point	Interval	Point	Interval
Cov. Adjustment	0.842	[0.692, 1.08]	0.184	[0.083, 0.286]
MLE	0.869	[0.711, 1.12]	0.166	[0.063, 0.269]
Nonlin. Method I	0.726	[0.690, 0.750]	0.155	[0.103, .278]
Nonlin. Method II	0.754	[0.710, 0.780]	0.144	[.119, .259]

Data of Solomon et al. (1981). $N = 14$.

TABLE 6. ESTIMATES OF TRUE BIOMASS ENERGETIC YIELD AND MAINTENANCE COEFFICIENT FROM BIOMASS, O_2, CO_2, NITROGEN AND GLUCOSE MEASUREMENT FOR THE GROWTH OF *CANDIDA UTILIS* ON UNFILTERED GLUCOSE FROM CORN DUST

Method of Analysis	η_{max}		m_e, h^{-1}	
	Point	Interval	Point	Interval
Cov. Adjustment	0.683	[0.567, 0.859]	0.115	[0.00, 0.231]
MLE	0.734	[0.564, 1.07]	0.128	[0.007, 0.249]
Nonlin. Method I	0.815	[0.749, 0.955]	0.019	[0.004, 0.026]
Nonlin. Method II	0.921	[0.855, 1.05]	0.098	[0.085, 0.104]

Data of Solomon et al. (1981). $N = 14$.

TABLE 7. POINT AND INTERVAL ESTIMATES OF TRUE ENERGETIC YIELD PARAMETERS OF CHEMOSTAT CULTURES OF *RIZOBIUM TRIFOLII**

	m_e, h^{-1}		η_{max}		ξ_p^{max}	
	Point	Interval	Point	Interval	Point	Interval
Equations are in Form I						
N. Lin. Regression of Eq. 7	0.02355	—	0.89966	—	0.45138	—
N. Lin. Regression of Eq. 8	0.00809	—	0.64943	—	0.73500	—
N. Lin. Method I	0.00715	[−0.00335, 0.01715]	0.63849	[0.45705, 0.80305]	0.77195	[0.52394, 1.12168]
N. Lin. Method II	0.00665	[−0.00093, 0.01425]	0.63306	[0.52832, 0.68386]	0.78166	[0.61453, 0.89365]
Cov. Adjustment	0.00658	[−0.01322, 0.02637]	0.63220	[0.47382, 0.94986]	0.78315	[0.50885, 1.69900]
Equations are in Form II						
N. Lin. Regression of Eq. 33	0.00989	—	0.70756	—	0.51924	—
N. Lin. Regression of Eq. 34	0.00410	—	0.59963	—	0.83510	—
N. Lin. Method I	0.00409	[−0.00670, 0.01329]	0.59789	[0.53422, 0.68911]	0.85395	[0.68143, 1.10871]
N. Lin. Method II	0.00401	[−0.00514, 0.01375]	0.59709	[0.52428, 0.70353]	0.85571	[0.67555, 1.14526]
Cov. Adjustment	0.00382	[−0.02017, 0.02782]	0.59529	[0.44524, 0.89793]	0.86011	[0.53638, 2.16943]

* Data of de Hollander et al. (1979). $N = 11$.

methods are quite appropriate as convergence is usually quick, however, as the number of responses increase, the covariance adjustment procedure is definitely preferred since it is very easy to use. The results from covariance adjustment may also be used as a starting point with the nonlinear methods.

From Table 7, independent estimates of the parameters by nonlinear regression of equations (Eqs. 7, 8, 33 and 34) show the need to simultaneously use all the information available. The use of the available electron and carbon balances in addition to the direct measurement of biomass and substrate data and when available nitrogen data, allow better point and interval estimates to be obtained.

Additional applications of the covariate adjustment approach are reported elsewhere, and an example is worked in considerable detail (Solomon et al., 1983).

ACKNOWLEDGMENT

This work was partly supported by NSF Grants CPE 79-18202 and CPE 81-20039.

NOTATION

A = $p \times N$ matrix of known coefficients and input variables in model (Eq. 21)
B = $p \times (p-1)$ matrix of rank $(p-1)$ chosen such that $B'\mathbf{1} = 0$
d = moles of carbon dioxide per g atom carbon of organic substrate, gmol/g atom carbon
E = $p \times N$ error matrix in Eq. 21
e_{jt} = element of error matrix
m = maintenance coefficient; m_e is g equivalent of available electrons of substrate consumed per g equivalent of biomass per hour
MLE = maximum likelihood estimate
N = number of observations
P = number of parameters, Eq. 47
p = number of observed responses in Eq. 21
Q_{CO_2} = specific rate of evolution of carbon dioxide, gmol/g dry wt. (h)
Q_{O_2} = specific rate of oxygen consumption, gmol/g dry wt. (h)
Q_s = specific rate of organic substrate consumption, g/g biomass (h)
q = number of covariates employed
r = number of parameters in Eq. 20
X = $p \times N$ response matrix
Y = biomass yield
$\mathbf{Y}$ = response vector of dimension N
$\bar{y}_{.i}$ = mean of p responses for the ith set of responses
$\hat{y}_{.i}$ = predicted value of response based on parameter estimates
y_c = biomass yield based on carbon, g biomass carbon per g substrate carbon
Z = covariate matrix of dimension $N \times q$ where $q = p - 1$ when all covariates are included
z = fraction of substrate carbon incorporated into products

Greek Letters

α_i = parameters associated with covariates
γ = reductance degree; γ_b is equivalents of available electrons of biomass per g atom carbon of biomass; $\gamma_b = 4 + p - 2n - 3q$ based on formula for biomass in Eq. 10 and the valences $C = 4$, $H = 1$, $O = -2$, and $N = -3$
ϵ = fraction of available electrons transferred to oxygen
ϵ_{iu} = element in error matrix
η = fraction of available electrons transferred to biomass; biomass energetic yield
η_{max} = true growth yield
μ = specific growth rate, h^{-1}
ξ = parameter vector in Eq. 21
ξ_p = fraction of available electrons incorporated into products; product energetic yield
ξ_p^{max} = true product yield
ρ = correlation coefficient
σ = mass fraction carbon
σ^2 = conditional variance associated with Eq. 22
$\hat{\sigma}^2$ = mean square error of conditional model

Subscripts

b = biomass
D = carbon dioxide
e = available electron basis
M = number of responses
N_2 = nitrogen
O_2 = oxygen
o = oxygen
p = product; number of responses
s = substrate

LITERATURE CITED

Ball, W. E., and L. C. D. Groenweghe, "Determination of Best-fit Rate Constants in Chemical Kinetics," *I&EC Fund.*, **5**, 181 (1966).

Box, G. E. P., and N. R. Draper, "The Bayesian estimation of common parameters from several responses," *Biometrika*, **52**, 355–365 (1965).

Brody, S., *Bioenergetics and Growth*, New York, Reinhold (1945).

Draper, N. R., and H. Smith, *Applied Regression Analysis*, Wiley, New York (1966).

Efron, B., "Bootstrap Methods: Another Look at the Jack Knife," *The Annals of Statistics*, **7**, No. 1, 1 (1979).

Efron, B., "Nonparametric Standard Errors and Confidence Intervals," Technical Rep. No. 67 Div. of Biostatistics, Stanford University, Stanford, CA (Apr., 1981).

Erickson, L. E., "Energetic Efficiency of Biomass and Product Formation," *Biotechnol. Bioeng.*, **21**, 725 (1979).

Erickson, L. E., "Growth and Product Energetic Yields of *Rhodopseudomonas sphaeroides* S in Dark and Aerobic Chemostat Cultures" *J. Fermentation Technol.*, **58**, 53 (1980).

Erickson, L. E., and J. L. Hess, "Analysis of Growth and Polysaccharide Yields in Chemostat Cultures of *Rhizobium trifolii*," *Annals N.Y. Acad. of Sci.*, **369**, 81 (1981).

Erickson, L. E., I. G. Minkevich, and V. K. Eroshin, "Application of Mass and Energy Balance Regularities in Fermentation," *Biotechnol. Bioeng.*, **20**, 1595 (1978).

Erickson, L. E., I. G. Minkevich, and V. K. Eroshin, "Utilization of Mass-Energy Balance Regularities in the Analysis of Continuous-Culture Data," *Biotechnol. Bioeng.*, **21**, 575 (1979).

Erickson, L. E., S. E. Selga, and U. E. Viesturs, "Application of Mass and Energy Balance Regularities in Fermentation," *Biotechnol. Bioeng.*, **20**, 1623 (1978).

Erickson, L. E., and S. Patel, "Applications of Available Electron Concepts and Regularities in Biological Heat Transfer and Thermodynamics," AIChE Annual Meeting Paper 105d, New Orleans, LA (1981).

Erickson, L. E., and S. A. Patel, "Analysis of Regularities and Process Efficiencies in Living Systems," *CRC Critical Reviews in Biomedical Engineering*, **8**, 311 (1982).

Ferrer, A., and L. E. Erickson, "Evaluation of Data Consistency and Estimation of Yield Parameters in Hydrocarbon Fermentations," *Biotechnol. and Bioeng.*, **21**, 2203 (1979).

Grizzle, J. E., and D. M. Allen, "Analysis of Growth and Dose Response Curves," *Biometrics*, **25**, 357 (1969).

Hempfling, W. P., and S. E. Mainzer, "Effects of Varying the Carbon Source Limiting Growth on Yield and Maintenance Characteristics of *Escherichia coli* in Continuous Culture," *J. Bacteriology*, **123**, 1076 (1975).

de Hollander, J. A., C. W. Bettenhoussen, and A. H. Stouthamer, "Growth Yields, Polysaccharide Production and Energy Conservation in Chemostat Culture of *Rhizobium trifolii*," *Antonie Van Leeuwenhock*, **45**, 401 (1979).

Hooke, R., and T. A. Jeeves, "Direct Search Solution of Numerical and Statistical Problem," *J. Assn. Comp. Mach.*, **8**, No. 2, 212 (Apr., 1961).

Hunter, W. G., "Estimation of Unknown Constants from Multiresponse Data," *I&EC Fund.*, **6**, 461 (1967).
Khatric, C. G., "A Note on a MANOVA Model Applied to Problems in Growth Curves," *Ann. Inst. Statist. Math.*, **18**, 75 (1966).
Meijer, E. M., H. W. Van Verseveld, E. G. Van der Beek, and A. H. Stouthamer, "Energy Conservation During Aerobic Growth of *Paracoccus dentrificans* in chemostat cultures," *Arch. Microbiol.*, **112**, 25 (1976).
Minkevich, I. G., and V. K. Eroshin, "Productivity and Heat Generation of Fermentation under Oxygen Limitation," *Filia Microbiologia*, **18**, 376 (1973).
Oner, M. D., L. E. Erickson, and S. S. Yang, "Estimation of True, Growth and Product Yields in Aerobic Cultures," *Biotechnol. Bioeng.*, **25**, 631 (1983).
Payne, W. J., "Energy Yields and Growth of Heterotrophs," *Annual Review of Microbiology*, **24**, 17 (1970).
Pirt, S. J., "The Maintenance Energy of Bacteria in Growing Cultures," *Proc. Royal Soc. London*, Ser. B., **163**, 224 (1965).
Pirt, S. J., *Principles of Microbe and Cell Cultivation*, Blackwell Scientific, London (1975).
Putthoff, R. R., and S. N. Roy, "A Generalized Multivariate Analysis of Variance Model Useful Especially for Growth Curves," *Biometrika*, **51**, 313 (1964).
Rao, C. R., "The Theory of Least Squares When the Parameters are Stochastic and Its Application to the Analysis of Growth Curves," *Biometrika*, **52**, 447 (1965).
Rao, C. R., "Covariance Adjustment and Related Problems in Multivariate Analysis," *Multivariate Analysis*, Academic Press, New York (1966).
Rao, C. R., "Least Square Theory Using an Estimated Dispersion Matrix and Its Application to Measurement of Signals," *Proc. Fifth Berkeley Symp. Math. Statis. Prob.*, **1**, 335 (1967).
SAS Users Guide, SAS Institute, Raleigh, NC (1979).
Solomon, B. O., L. E. Erickson, and J. L. Hess, "Applications of Data Consistency Tests and New Parameter Estimation Methods for Microbial Growth on Corn Dust in Batch Culture," *Biotechnol. Bioeng.*, **23**, 2333 (1981).
Solomon, B. O., L. E. Erickson, J. L. Hess, and S. S. Yang, "Maximum Likelihood Estimation of Growth Yields," *Biotechnol. Bioeng.*, **24**, 633 (1982).
Solomon, B. O., L. E. Erickson, and S. S. Yang, "Utilization of Statistics and Experimental Design in Data Collection and Analysis," *Biotechnol. Bioeng.*, **25**, 2683 (1983).
Van Verseveld, H. W., and A. H. Stouthamer, "Oxidative Phosphorylations of *Micrococcus Dentrificans*—Calculation of the P/O Ratio in Growing Cells," *Arch. Microbiol.*, **107**, 241 (1976).
Yang, S. S., B. O. Solomon, M. D. Oner, and L. E. Erickson, "A Method of Estimation and Testing Common Parameters for Some Multiresponse Models Associated with Growth and Bioenergetics," *Technometrics*, (1984).

Manuscript received June 29, 1982; revision received September 1, and accepted September 7, 1983.

Mass Transfer in AC Electrolysis

Part I: Theoretical Analysis Using a Film Model for Sinusoidal Current on a Rotating Hemispherical Electrode

C. Y. CHENG and D-T. CHIN

Department of Chemical Engineering
Clarkson University
Potsdam, NY 13676

A film model is presented for the analysis of mass transfer to a rotating hemispherical electrode when sinusoidal alternating current (AC) together with direct current (DC) are flowing across the electrode surface. The concentration of a diffusing ion is separated into two independent components: a constant DC component and a periodic AC component. The DC concentration is obtained by solving the steady-state convective mass transport equation with the perturbation method. The periodic AC concentration distribution is analyzed by the solution to the one-dimensional transient diffusion equation based on the concept of Nernst diffusion layer. The limiting AC current densities corresponding to a zero surface concentration of a reactive ion are investigated for various DC current densities and AC frequencies. The resulting periodic concentration overpotential wave and its phase shift with respect to the applied AC are examined. A comparison with a previous rigorous model indicates that the film model is a good approximation to the mass transfer calculation in the regimes of a dimensionless AC frequency $K = (\omega/\Omega)Sc^{1/3}$ greater than 2 and less than 0.01.

SCOPE

Electrolysis with a direct current superimposed with a periodically alternating current component has been long used in pulse plating, in electrodissolution to increase anode corrosion, in AC anodizing to improve color and dye penetration, and in the AC corrosion processes (Venkatesh, 1979). AC can enhance the kinetics of electrochemical reactions, change the morphology of electrodeposits, and cause the pitting corrosion of passive metals. AC also improves the mass transfer rate by producing a pulsating concentration boundary layer of reactive ions near the electrode surface, and a very large AC of an order of 10–100 times greater than the DC-limitng current density can be used for electrolysis without affecting the coulombic efficiency of the electrode reaction.

This work examines the mass transfer to a rotating hemi-