2002

Cell Population Modeling and Parameter Estimation for Continuous Cultures of Saccharomyces cerevisiae

Prashant Mhaskar
Martin A. Hjortse
Michael A Henson, *University of Massachusetts - Amherst*

Cell Population Modeling and Parameter Estimation for Continuous Cultures of *Saccharomyces cerevisiae*

Prashant Mhaskar and Martin A. Hjortsø

Department of Chemical Engineering, Lousiana State University, Baton Rouge, Louisiana 70803-7303

Michael A. Henson*

Department of Chemical Engineering, University of Massachusetts, Amherst, Massachusetts 01003-3110

Saccharomyces cerevisiae is known to exhibit sustained oscillations in chemostats operated under aerobic and glucose-limited growth conditions. The oscillations are reflected both in intracellular and extracellular measurements. Our recent work has shown that unstructured cell population balance models are capable of generating sustained oscillations over an experimentally meaningful range of dilution rates. A disadvantage of such unstructured models is that they lack variables that can be compared directly to easily measured extracellular variables. Thus far, most of our work in model development has been aimed at achieving qualitative agreement with experimental data. In this paper, a segregated model with a simple structured description of the extracellular environment is developed and evaluated. The model accounts for the three most important metabolic pathways involved in cell growth with glucose substrate. As compared to completely unstructured models, the major advantage of the proposed model is that predictions of extracellular variables can be compared directly to experimental data. Consequently, the model structure is well suited for the application of estimation techniques aimed at determining unknown model parameters from available extracellular measurements. A steady-state parameter selection method developed in our group is extended to oscillatory dynamics to determine the parameters that can be estimated most reliably. The chosen parameters are estimated by solving a nonlinear programming problem formulated to minimize the difference between predictions and measurements of the extracellular variables. The efficiency of the parameter estimation scheme is demonstrated using simulated and experimental data.

1. Introduction

Saccharomyces cerevisiae is an essential microorganism in industries such as brewing, baking, and food manufacturing. Yeasts such as *S. cerevisiae* also are important in genetic engineering applications. Under certain environmental conditions, continuous cultures of *S. cerevisiae* exhibit sustained oscillations in extracellular variables such as carbon dioxide evolution rate; oxygen uptake rate; glucose, ethanol, and acetic acid concentrations; pH controller action; and intracellular variables such as biomass concentration, cell size distribution, protein content, and storage carbohydrate concentrations (*1*−*12*). Three types of oscillations have been reported in the literature (*13*, *14*): cell cycle dependent oscillations, glycolytic oscillations and short-period sustained oscillations. A marked cell cycle synchronization of the yeast population is observed in association with cell cycle dependent oscillations, and strong interactions between synchronized cell cycle events and oscillating metabolic signals have been reported (*5*, *7*, *8*, *10*−*12*). Cell cycle dependent oscillations are observed under aerobic growth conditions and in glucose limited environments. Oscillations appear at intermediate dissolved oxygen levels (*3*, *5*, *10*) and for a wide range of dilution rates (typically 0.09−0.25 h^{-1}) (*9*). They are particularly difficult to suppress in high performance chemostats at dilution rates between 0.13 and 0.2 h^{-1} (*11*). The period of oscillation varies from 2 to 45 h depending on the strain and culture conditions (*3*).

A number of mathematical models have been proposed to rationalize the sustained oscillations observed in continuous cultures of *S. cerevisiae*. Existing models can be classified into three general categories: structured and unsegregated (*15*), unstructured and segregated (*16*), and structured and segregated (*11*). Structured unsegregated models are based on the assumption that sustained oscillations are attributable to cell metabolism rather than cell population dynamics. A shortcoming of such models is that they cannot adequately explain cell cycle synchronization, which plays a critical role in oscillatory yeast dynamics (*8*). Unstructured segregated models are based on the assumption that oscillations arise as a result of interactions between the cell population and the extracellular environment. A key feature of these models is the population balance equation (PBE) that describes the time evolution of the distribution of states (e.g., cell age or cell mass distributions). While they are capable

* To whom correspondence should be addressed: Ph: 413-545-3481. Fax: 413-545-1647. E-mail: henson@ecs.umass.edu.

10.1021/bp020083i CCC: $22.00
Published on Web 07/30/2002

of predicting cell cycle synchrony, these models cannot capture the interplay between cell metabolism and oscillatory dynamics. Structured segregated models have been developed to address the limitations of the simpler models described above. While such models provide the most rigorous description of cell growth, they are mathematically complex and require the availability of sophisticated intracellular measurement technology such as flow cytometry (*17*) for validation.

Although unstructured segregated models can capture the qualitative behavior of cell growth, the absence of easily measured variables renders these models unsuitable for direct comparison with experimental data. On the other hand, these models are easier to formulate and more computationally efficient than structured segregated models. In this paper, we pursue a different modeling approach in which an unstructured cell population model is coupled to a simple structured description of the extracellular environment. The extracellular model is structured in the same sense that intracellular models are structured: multiple reaction products and their interactions in the extracellular environment are captured. Clearly the resulting model is not a substitute for segregated yeast models that include a detailed description of the intracellular reactions. Rather the structured medium description is an extension that allows model predictions to be directly compared with available extracellular measurements.

Our previous modeling work has been aimed at capturing qualitative dynamic behavior that is observed experimentally. The proposed model provides an opportunity to systematically estimate key model parameters to achieve quantitative agreement with experimental data. The model structure and the type of extracellular measurements limit the number of parameters that can be estimated reliably. We present a method that utilizes the available model to determine the parameters that are most readily estimated from a particular set of steady-state or oscillatory measurements. Based on these results, estimates of the model parameters are found by solving a suitably formulated nonlinear programming problem.

The remainder of the paper is organized as follows: In section 2, the cell population model with structured medium is described. In section 3, a steady-state parameter selection procedure previously developed in our group is extended to oscillatory data sets. Parameter estimation results using simulated and experimental data also are presented. Section 4 summarizes the paper and presents the major conclusions.

2. Cell Population Model with Structured Medium

The proposed model consists of a segregated description of the cell population and a simple structured description of the growth medium. This section focuses on development of the dynamic equations governing the extracellular environment. The segregated cell model is outlined below with an emphasis on the modifications required to accommodate the structured medium description. Our previous work (*18*) should be consulted for details on the segregated yeast model.

2.1. Cell Population Model. Cell division in the budding yeast cell cycle is asymmetric. The smaller of the newborn cells generated by division is referred to as a daughter cell and the larger cell is called a mother cell. Newborn daughter cells must grow to attain the size characteristic of newborn mother cells (characterized here by the cell transition mass) before starting a budding cycle, while newborn mother cells bud shortly after being born. After budding has occurred, the bud grows while the mass of the mother cell remains essentially constant. Cell division (characterized here by the cell division mass) produces a newborn daughter cell and a newborn mother cell that subsequently progress through the cell cycle. Based on this simple cell cycle model, a population balance equation (PBE) that describes the evolution of the cell mass distribution is formulated as follows (*19*):

$$\frac{\partial W(m,t)}{\partial t} + \frac{\partial [K(S')W(m,t)]}{\partial m} = \int_0^{m'} 2p(m,m')\Gamma(m',S')W(m',t)\,dm' - [D + \Gamma(m)]W(m,t) \quad (1)$$

where m is the cell mass; $W(m, t)$ is the distribution of cell mass; $K(S')$ is the overall single cell growth rate, which is the sum of the growth rates due to three metabolic pathways introduced below; S' represents the intracellular substrate concentration, which is the sum of the intracellular concentrations of glucose and ethanol (explained later); $p(m, m')$ is the newborn cell mass distribution function; $\Gamma(m, S')$ is the division intensity function; and D is the dilution rate.

The division intensity function is introduced to account for the probabilistic nature of cell division:

$$\Gamma(m,S') = \begin{cases} 0 & m \le m_t^* + m_o \\ \gamma e^{-\epsilon(m - m_d^*)^2} & m \in [m_t^* + m_o,\ m_d^*] \\ \gamma & m \ge m_d^* \end{cases} \quad (2)$$

where m_t^* is the cell transition mass, m_o is the additional mass beyond m_t^* that mother cells must attain before division is possible, m_d^* is the division mass, and ϵ and γ are constant parameters. The newborn cell mass distribution function has the form

$$p(m,m') = A\exp[-\beta(m - m_t^*)^2] + A\exp[-\beta(m - m' + m_t^*)^2] \quad (3)$$

when $m < m'$ and $m' > m_t^* + m_o$; the function is identically zero otherwise. Here A and β are constant parameters. This function yields two Gaussian peaks in the cell number distribution, one centered at m_t^* corresponding to mother cells and one centered at $m_t^* - m'$ corresponding to daughter cells.

Sustained oscillations are generated through an induction synchrony mechanism (*20*) in which the transition and division masses are functions of the nutrient concentration. The following saturation functions are used:

$$m_t^*(S') = \begin{cases} m_{to} + K_t(S_l - S_h) & S' < S_l \\ m_{to} + K_t(S' - S_h) & S' \in [S_l,\ S_h] \\ m_{to} & S' > S_h \end{cases} \quad (4)$$

$$m_d^*(S') = \begin{cases} m_{do} + K_d(S_l - S_h) & S' < S_l \\ m_{do} + K_d(S' - S_h) & S' \in [S_l,\ S_h] \\ m_{do} & S' > S_h \end{cases} \quad (5)$$

where S_l, S_h, m_{to}, m_{do}, K_t and K_d are constants. For the nominal parameter values used subsequently, these functions are consistent with experimental observations that the ratio of the division mass to the transition mass increases with increasing nutrient concentration (*1, 5*).

2.2. Structured Medium Model. Oscillatory yeast dynamics are observed in glucose-limited growth environments. Under such conditions, both glucose and the excreted product ethanol can serve as substrates for cell growth. The following reaction sequence accounts for the relevant metabolic pathways, glucose fermentation, glucose oxidation, and ethanol oxidation:

$$C_6H_{12}O_6 \rightarrow 2C_2H_5OH + 2CO_2 \quad (6)$$

$$C_6H_{12}O_6 + 6O_2 \rightarrow 6CO_2 + 6H_2O \quad (7)$$

$$C_2H_5OH + 3O_2 \rightarrow 2CO_2 + 3H_2O \quad (8)$$

The glucose fermentation rate (K_{gf}) is assumed to follow Monod kinetics with respect to the substrate glucose. The glucose oxidation rate (K_{go}) and the ethanol oxidation rate (K_{eo}) are assumed to follow Monod kinetics with respect to both substrate and dissolved oxygen. Experiments show that glucose is preferentially consumed when both substrates are available (*21*). This effect is modeled by introducing a glucose inhibitory term in the ethanol oxidation rate. The resulting rate expressions are

$$K_{gf}(G') = \frac{\mu_{mgf} G'}{K_{mgf} + G'} \quad (9)$$

$$K_{go}(G', O) = \frac{\mu_{mgo} G'}{K_{mgo} + G'} \frac{O}{K_{mgd} + O} \quad (10)$$

$$K_{eo}(G', E', O) = \frac{\mu_{meo} E'}{K_{meo} + E'} \frac{O}{K_{med} + O} \frac{K_{inhib}}{K_{inhib} + G'} \quad (11)$$

where G' and E' represent intracellular glucose and ethanol concentrations, respectively; O is the dissolved oxygen concentration; μ_{mgf}, μ_{mgo}, and μ_{meo} are maximum consumption rates; K_{mgf}, K_{mgo}, K_{mgd}, K_{meo} and K_{med} are saturation constants; and K_{inhib} is a constant that characterizes the inhibitory effect of glucose on ethanol oxidation. Note that the rate expressions do not account for the so-called Crabtree effect, which causes repression of glucose oxidation under high glucose concentrations commonly observed during batch growth (*11*). While the Crabtree effect is easily incorporated into the model, there is little motivation to introduce this additional complication since the focus of this study is oscillatory dynamics where the glucose concentration is comparatively low.

The substrate mass balance equations are

$$\frac{dG}{dt} = D(G_f - G) - \int_0^\infty \left[\frac{K_{gf}(G')}{Y_{gf}} + \frac{K_{go}(G')}{Y_{go}}\right] W(m, t)\, dm \quad (12)$$

$$\frac{dE}{dt} = D(E_f - E) + \frac{92}{180} \int_0^\infty f(m) \frac{K_{gf}(G')}{Y_{gf}} W(m, t)\, dm - \int_0^\infty \frac{K_{eo}(E')}{Y_{eo}} W(m, t)\, dm \quad (13)$$

where G and E are the extracellular glucose and ethanol concentrations, respectively; G_f and E_f are the feed glucose and ethanol concentrations, respectively; Y_{gf}, Y_{go}, and Y_{eo} are constant yield coefficients; and the ratio 92/180 represents the mass of ethanol produced per mass of glucose fermentatively consumed. Experimental data suggests that key products such as ethanol are excreted primarily by budded cells (*1*, *22*). The function $f(m)$ in eq 13 is used to model this behavior:

$$f(m) = \begin{cases} 0 & m \le m_t^* \\ \gamma_e \exp[-\epsilon_e (m - m_t^* - m_e)^2] & m > m_t^* \end{cases} \quad (14)$$

where γ_e, ϵ_e, and m_e are constant parameters.

The transport of substrates from the extracellular environment into the yeast cells does not occur instantaneously. Rather than utilize rather complex mechanistic models (*23*), we have chosen to pursue a simpler phenomenological approach in which the uptake of each substate is modeled as a first-order process:

$$\frac{dG'}{dt} = \alpha_g (G - G') \quad (15)$$

$$\frac{dE'}{dt} = \alpha_e (E - E') \quad (16)$$

where G and E are extracellular concentrations; G' and E' are intracellular concentrations; and α_g and α_e are time constants for the glucose and ethanol substrate uptake processes, respectively. While they are included to make the dynamic model more realistic, these equations are not necessary to generate sustained oscillations.

The liquid-phase oxygen balance is written as

$$\frac{dO}{dt} = K_{lo} a (O^* - O) - \int_0^\infty \left[\frac{192}{180} \frac{K_{go}(G')}{Y_{go}} + \frac{96}{46} \frac{K_{eo}(E')}{Y_{eo}}\right] W(m, t)\, dm \quad (17)$$

where O^* is the saturation oxygen concentration; K_{lo} is the oxygen mass transfer coefficient; a is the interfacial area per unit liquid volume; the ratio 192/180 represents the mass of oxygen consumed per mass of glucose oxidatively metabolized; and the ratio 96/46 represents the mass of oxygen consumed per mass of ethanol metabolized. The oxygen solubility is assumed to be governed by Henry's law:

$$O^* = H_o R T O_{out} \quad (18)$$

where H_o is the Henry's rate constant for oxygen; O_{out} is oxygen partial pressure in the gas exhaust stream; T is the absolute temperature; and R is the gas constant. The gas-phase oxygen balance is

$$\frac{dV_g O_{out}}{dt} = F(O_{in} - O_{out}) - K_{lo} a (O^* - O) V_l \quad (19)$$

where V_g and V_l are the gas phase and liquid-phase volumes, respectively; F is the volumetric air feed flow rate; and O_{in} is the oxygen partial pressure in the air feed stream.

The liquid-phase carbon dioxide balance is

$$\frac{dC}{dt} = K_{lc} a (C^* - C) + \int_0^\infty \left[\frac{264}{180} \frac{K_{go}(G')}{Y_{go}} + \frac{88}{46} \frac{K_{eo}(E')}{Y_{eo}}\right] W(m, t)\, dm + \int_0^\infty \left[f(m) \frac{88}{180} \frac{K_{gf}(G')}{Y_{gf}}\right] W(m, t)\, dm \quad (20)$$

where C is the liquid-phase carbon dioxide concentration; C^* is the saturation carbon dioxide concentration; K_{lc} is

the carbon dioxide mass transfer coefficient; the ratio 264/180 represents the mass of carbon dioxide produced per mass of glucose oxidatively metabolized; the ratio 88/46 represents the mass of carbon dioxide produced per mass of ethanol metabolized; and the ratio 88/180 represents the mass of carbon dioxide produced per mass of glucose fermentatively metabolized. The carbon dioxide solubility is modeled as

$$C^* = H_c(\mathrm{pH})RTC_{out} \tag{21}$$

where H_c is the Henry's constant for carbon dioxide evaluated at a pH of 5.0; and C_{out} is the carbon dioxide partial pressure in the exhaust gas stream. The gas-phase carbon dioxide balance is

$$\frac{dV_gC_{out}}{dt} = F(C_{in} - C_{out}) - K_{lc}a(C^* - C)V_l \tag{22}$$

where C_{in} is the carbon dioxide partial pressure in the air feed stream.

2.3. Model Parameters and Numerical Solution. The dynamic yeast model contains 41 parameters that must be specified. Values of parameters such as yield coefficients and maximum growth rates are available in the literature (*21*, *24*). Glucose saturation constants are chosen such that glucose fermentation occurs primarily when the glucose concentration is high ($\gg$1 g/L) and glucose oxidation predominates at low glucose concentrations ($\approx$1 g/L). It has been observed that ethanol oxidation occurs primarily when the glucose concentration is lower than 0.1 g/L (*21*). Therefore K_{inhib} is chosen such that the effect of glucose inhibition on the ethanol oxidation rate is prominent at concentrations higher than 0.1 g/L. Values of operating parameters such as the dilution rate and air flow rate are chosen to match the experimental conditions used in our laboratory (*25*). The "known" parameters and their values are listed in Table 1. Unfortunately, there is no simple basis to select the parameter values in the segregated model functions. Table 2 shows nominal values for these "unknown" parameters, which have been chosen to obtain qualitative agreement between the model and experimental data with respect to the period, amplitudes, and dilution rate range of the oscillatory signals.

The proposed dynamic model consists of a partial differential-integro equation for the cell mass distribution coupled to ordinary differential-integro equations representing the extracellular environment. We have found that orthogonal collocation on finite elements allows efficient and robust solution of such cell population models (*18*). The mass domain is discretized such that the population balance equation is approximated by a set of nonlinear ordinary differential equations in time. Integral expressions are approximated using Gaussian quadrature. The resulting nonlinear ordinary differential equation model is integrated using the MATLAB code ode15s.

3. Model Parameter Estimation

The parameter values in Table 2 were adjusted by hand to achieve qualitative agreement with experimental observations. The problem of determining "optimal" parameter values is nontrivial since their effect on model behavior is complex. Because the proposed model provides predictions of readily measurable extracellular variables such as dissolved oxygen and evolved carbon dioxide concentrations, experimental data can be used to estimate unknown model parameters in an attempt to achieve quantitative agreement. In section 3.1, a model-based method for determining which unknown parameters can be estimated most reliably for a given set of oscillatory measurements is presented. An optimization based strategy for estimating the chosen parameters from dynamic data is discussed in section 3.2.

Table 1. Experimentally Derived Model Parameters

variable	value	variable	value
Y_{gf}	0.15 g/g	μ_{mgf}	30×10^{-13} g/h
Y_{go}	0.65 g/g	μ_{mgo}	3.25×10^{-13} g/h
Y_{eo}	0.5 g/g	μ_{meo}	7×10^{-13} g/h
K_{mgf}	40 g/l	K_{mgo}	2 g/L
K_{meo}	1.3 g/L	K_{inhib}	0.4 g/L
K_{mgd}	0.001 g/L	K_{med}	0.001 g/L
H_o	0.0404 g/L/atm	H_c	1.48 g/L/atm
V_g	0.9 l	V_l	0.1 l
K_la	1500 h^{-1}	K_lc	1500 h^{-1}
D	0.15 h^{-1}	G_f	30 g/L
E_f	0 g/L	F	90 L/h
T	298 K	O_{in}	0.21 atm
C_{in}	0.0003 atm		

Table 2. Other Model Parameters

variable	value	variable	value
γ	400	ϵ	7
γ_e	8	ϵ_e	20
m_o	1×10^{-13} g	m_e	1.54×10^{-13} g
A	$\sqrt{(10/\pi)}$	β	40
S_l	0.1 g/L	S_h	2.0 g/L
K_t	0.01 g/g·L	K_d	3.83 g/g·L
m_{to}	4.55×10^{-13} g	m_{do}	10.25×10^{-13} g
α_g	20	α_e	20

3.1. Parameter Selection. The model parameters best suited for estimation are determined by the model structure, the available measurements and the experimental data. Below we present a parameter selection method in which the first two factors are considered explicitly. As discussed below, the effect of the available experimental data must be evaluated numerically. The objective is to select a subset of the unknown model parameters that have the strongest influence on model behavior. The discretized yeast model is represented as

$$\dot{\hat{x}} = f(\hat{x}, u, \theta_{fix}, \theta_{est}) \tag{23}$$

$$\hat{y} = h(\hat{x}, u, \theta_{fix}, \theta_{est}) \tag{24}$$

where $\hat{x}$ is the state vector, which consists of cell number distribution values at the collocation points and the extracellular variables; $\hat{y}$ is the vector of measured output variables to be defined below; u is a vector of input variables; θ_{fix} is a vector containing the "known" model parameters listed in Table 1 that are fixed at constant values and not used for estimation; and θ_{est} is a vector containing the "unknown" parameters listed in Table 2 that are available for estimation.

Two issues must be considered in selecting the model parameters to be estimated. First, the parameters should have a strong effect on the output measurements. Otherwise large changes in parameter values are required to fit the data. Second, the effect of each parameter on the measured outputs should be unique. Parameters whose effects are strongly collinear lead to nonunique values when estimation is performed with limited plant data. The parameters chosen for estimation should provide an appropriate balance between these two objectives. Below we introduce the concept of "parameter identifiability" as a measure of the potential to uniquely determine a particular parameter from a given set of measurements.

While other methods are available (*26*−*28*) we utilize a parameter selection procedure developed in our group (*29*) that accounts for output sensitivities to parameter variations and collinearities between parameter effects. First the procedure is described for steady-state analysis. The model is used to compute steady-state output values for nominal and perturbed values of each parameter in the set θ_{est}. Elements of the parameter-output sensitivity matrix K are calculated as

$$K_{ij} = \frac{\Delta\hat{\bar{y}}_i/\hat{\bar{y}}_{n,i}}{\Delta\theta_{est,j}/\theta_{est,j}} \tag{25}$$

where $\theta_{est,j}$ is the nominal value of the j-th adjustable parameter; $\Delta\theta_{est,j}$ is the difference between the perturbed and nominal values of the j-th adjustable parameter; $\hat{\bar{y}}_{ni}$ is the steady-state value of the i-th measured output obtained with the nominal parameter values; and $\Delta\hat{\bar{y}}_i$ is the difference between the perturbed and nominal steady-state values of the i-th measured output.

Principle component analysis (PCA) (*30*, *31*) is performed on the sensitivity matrix to evaluate the overall effect of each parameter on the measured outputs. The principle components are the eigenvectors of the covariance matrix KK^T. The weighted sum of the principal components and their associated eigenvalues is a measure of the overall parameter effect:

$$w_j = \sum_{i=1}^{m} |\lambda_i P_{ij}| \tag{26}$$

where m is the number of measurements; P_{ij} is the j-th element of the i-th principle component; and λ_i is the i-th eigenvalue. Absolute values are used because the sign of each individual effect is not relevant. The highest ranked parameter has the largest overall effect w_j. The remaining parameters are ranked to provide a suitable compromise between their overall effect and their collinearity with higher ranked parameters. The sine of the angle between the sensitivity vectors of two parameters is a measure of their collinearity:

$$\sin \alpha_{\theta_i,\theta_j} = \frac{\langle K_i, K_j \rangle}{||K_i||\cdot||K_j||} \tag{27}$$

where K_i is the sensitivity vector of parameter θ_i; $< \cdot,\cdot >$ denotes the inner product; $||\cdot||$ denotes the 2-norm; and $\alpha \in [0, \pi]$. If the sensitivity vectors are orthogonal, then $\sin \alpha_{\theta_i,\theta_j} = 1$ and the effects of the two parameter are easily distinguished. Conversely, $\sin \alpha_{\theta_i,\theta_j} = 0$ when the sensitivity vectors are parallel and the parameter effects cannot be distinguished.

The identifiability of a particular parameter is defined as its overall effect weighted by its collinearity with higher ranked parameters:

$$J_{N,i} = w_i \prod_{j=1}^{N} \sin \alpha_{\theta_j,\theta_i}; \quad i = 1, ..., n; i \neq j \tag{28}$$

where n is the number of parameters; and $J_{N,i}$ is the identifiability of the i-th parameter when N parameters are ranked higher. These calculations can be used iteratively to generate a ranking of adjustable parameters (*29*).

As described above, the parameter selection procedure is suitable only for steady-state analysis. Consequently, this method may rank highly parameters that strongly affect steady-state behavior but have little effect on dynamic behavior such as sustained oscillations. The parameter ranking procedure can be extended to limit cycle behavior by characterizing the periodic solution with the amplitude and mean value of each measured variable as well as the oscillation period. Then the sensitivity of a given parameter is defined as the normalized difference between the mean, amplitude and period of fully developed oscillations for perturbed and nominal values of the parameter. Based on this modified sensitivity matrix, the same iterative procedure can be used to determine parameter rankings which are better suited for fitting oscillatory data.

The ranking determines which parameters should be estimated given that the number of estimated parameters has been chosen. The appropriate number of estimated parameters is affected by the quantity and quality of the available data. If the number of parameters selected is too small, the model will produce unsatisfactory predictions. The use of too many parameters leads to nonuniqueness of the estimates for different initial parameter values. Both situations are undesirable and should be avoided. Below we show that the appropriate number of estimated parameters can be determined by limited simulation studies.

Because the sensitivity matrix is constructed by perturbing a nominal steady-state or oscillatory solution, the analysis is inherently local. We have found that the parameter ranking usually is weakly dependent on the perturbation as long as the perturbation magnitude is sufficiently small. By contrast, the nominal parameter values can have a strong effect on the ranking procedure. This emphasizes the need to specify "good" initial values for the adjustable parameters and to exercise engineering judgment in interpreting the results.

In the subsequent simulation studies, the measured variables are assumed to be percent carbon dioxide and percent oxygen in the gas exhaust stream, the percent dissolved oxygen, and the glucose, ethanol, and cell number concentrations in the liquid phase. Steady-state data for parameter sensitivity analysis are generated at a dilution rate of 0.1 h^{-1} and an air flow rate of 90 L/h. The "unknown" parameters in Table 2 are perturbed by 1% from their nominal values, while the "known" parameters in Table 1 remain fixed at their nominal values. In the dynamic case, sustained oscillations at a dilution rate of 0.15 h^{-1} and an air flow rate of 90 L/h represent the nominal state. The "unknown" parameters are varied by one percent to generate perturbed limit cycles.

Results of the parameter ranking procedure for steady-state analysis are shown in the first column of Table 3. Saturation values of the division mass (m_{do}) and the transition mass (m_{to}) are ranked first and second, respectively. Because they determine the cell mass that mother and daughter cells attain, these two parameters have strong effects on the amount of substrate consumed and the output signals produced. The parameter m_e ranked third has the largest effect on the steady-state ethanol concentration. Because the nominal substrate concentration is much greater than the lower limit in the transition and division mass functions, S_l is identified as the third least important parameter. The filter parameters (α_g, α_e) are ranked last because they have no steady-state effect on the measurements.

Results of the parameter ranking procedure for dynamic analysis are shown in the second column of Table 3. As before, m_{do} is ranked as the most important parameter. The ethanol production parameter m_e is chosen as the second most important parameter, and the

Table 3. Parameter Rankings

simulation steady-state		simulation dynamic		experimental dynamic	
parameter	J_i	parameter	J_i	parameter	J_i
m_{do}	9.1×10^{-1}	m_{do}	9.2×10^{-1}	m_{do}	8.2×10^{-1}
m_{to}	3.9×10^{-1}	m_e	3.5×10^{-1}	m_e	3.3×10^{-1}
m_e	9.2×10^{-2}	K_t	1.2×10^{-1}	m_{to}	1.5×10^{-1}
γ_e	1.1×10^{-3}	K_d	3.4×10^{-2}	γ_e	4.31×10^{-2}
S_h	8.5×10^{-4}	m_{to}	2.2×10^{-2}	K_d	3.6×10^{-2}
ϵ	2.7×10^{-4}	ϵ	3.6×10^{-3}	ϵ_e	1.75×10^{-2}
ϵ_e	4.9×10^{-5}	m_o	1.5×10^{-3}	ϵ	9.7×10^{-3}
K_d	1.8×10^{-5}	γ	1.2×10^{-3}	S_l	9.85×10^{-4}
γ	7.2×10^{-6}	A	1.1×10^{-3}	γ	5.36×10^{-5}
K_t	6.7×10^{-6}	S_l	9.5×10^{-4}	A	3.69×10^{-6}
A	9.4×10^{-7}	S_h	4.0×10^{-4}	K_t	1.21×10^{-6}
m_o	6.1×10^{-7}	γ_e	6.0×10^{-5}	m_o	3.48×10^{-8}
S_l	5.4×10^{-8}	ϵ_e	3.2×10^{-6}	S_h	1.09×10^{-8}
α_g,α_e	3.8×10^{-13}	α_g,α_e	1.9×10^{-8}	α_g,α_e	1.6×10^{-9}

slope of the division mass function K_t is ranked third. These results can be explained by noting that oscillations are attributable to synchrony in the cell mass distribution. The parameter m_e determines the overlap of the ethanol production function with the cell mass distribution, thereby determining the amount of ethanol excreted during each cell cycle. The slope K_t determines the values between which the transition mass m_t^* varies. Therefore, this parameter has stronger effect on oscillation amplitudes than does the saturation value m_{to}. Note that the parameter A in the newborn probability distribution function is considered to be much more important than in the steady-state case. This can be explained by noting that A determines the dispersive effect of the probability function, thereby influencing the oscillatory behavior. The effect of the ethanol production function slope ϵ_e is found to be highly collinear with that of m_e; therefore ϵ_e is ranked as the third least important parameter. The parameters α_g and α_e are ranked last because they have little influence on the means, amplitudes and period of the sustained oscillations.

In our laboratory, the glucose and cell number concentrations must be measured off-line. These measurements were unavailable at the time experimental data was collected. Furthermore, the on-line measurement of liquid-phase ethanol concentration was not sufficiently sensitive to provide reliable data during sustained oscillations. Therefore, the ranking procedure was repeated with these three measurements excluded from the set of available measurements. Results for dynamic analysis are shown in the third column of Table 3. The first two parameters are the same as those obtained when the additional three measurements are available (second column of Table 3). The parameter previously ranked third (K_t) is unimportant as a result of the lack of the cell number concentration measurement. Instead, m_{to} is ranked third because it has a relatively large effect on the mean values of the available measurements.

3.2. Parameter Estimation from Simulated Steady-State Data. The parameter selection method yields a ranking of the adjustable parameters in terms of their identifiability. The number of parameters that can be estimated reliably is not known a priori from the ranking. For this purpose, limited simulation tests are conducted to determine the number of parameters that can be estimated from the available data. The primary motivation for such simulation tests is to provide the necessary analysis for the parameter selection and estimation methods. Otherwise, parameter estimates subsequently derived from experimental data will be rather dubious.

Table 4. Plant and Perturbed Parameters θ_2 for Steady-State Analysis

		parameters			
parameter	plant value	1	2	3	4
m_{do}	10.25				
m_{to}	4.55	5.25			
m_e	1.54	2.54	2.54		
γ_e	1.25	1.55	1.55	1.55	
S_h	2	2.3	2.3	2.3	2.3
ϵ_e	20	27	27	27	27
ϵ	7	10.94	10.94	10.94	10.94
K_d	3.83	3.95	3.95	3.95	3.95
γ	400	136.78	136.78	136.78	136.78
K_t	0.01	0.021	0.021	0.021	0.021
A	10	6.77	6.77	6.77	6.77
m_o	1	0.65	0.65	0.65	0.65
S_l	0.1	0.2	0.2	0.2	0.2
α_g	20	27.23	27.23	27.23	27.23
α_e	20	15.11	15.11	15.11	15.11

Table 5. Initial and Final Parameters θ_1 Estimated from Steady-State Data

		parameters					
		1		2		3	
parameter	plant	initial	final	initial	final	initial	final
m_{do}	10.25	10.5	10.89	10.5	10.89	11.1	10.68
m_{to}	4.55			4.75	4.97	5	4.58
m_e	1.54					1.8	1.63

Values of the parameters θ_{fix} listed in Table 1 are assumed to be known. The model with the nominal values for the parameters θ_{est} listed in Table 2 is termed the "plant" and is used to generate steady-state simulation data at different operating conditions. The unknown parameter set θ_{est} is divided into two subsets. The first subset θ_1 consists of the unknown parameters that are actually estimated, while the second subset θ_2 contains the remaining unknown parameters that are not estimated. Parameters in θ_1 are selected on the basis of the ranking generated by the parameter selection method. Only the parameters θ_1 are used as decision variables in the optimization-based parameter estimation strategy described below.

To test the parameter estimation method, initial values of the unknown parameters θ_1 used in the model are chosen to be significantly different than the values listed in Table 2. Values of the other unknown parameters θ_2 used in the model are obtained by randomly perturbing the plant values listed in Table 2 using a Gaussian distribution with zero mean and variance of 0.3. Table 4 shows values of the plant and perturbed parameters θ_2 for 1−4 estimated parameters. The objective is to uniquely determine values of the parameters θ_1 such that the error between the model predictions and plant outputs is minimized. Because the unknown parameters θ_2 are perturbed from their nominal plant values, the estimated values of the parameters θ_1 are not expected to converge to the plant values.

Steady-state values of the dissolved oxygen, evolved carbon dioxide, evolved oxygen, ethanol, glucose, and total cell number concentrations at different air flow rates and dilution rates are used as the data set for parameter estimation. Data are generated at two air flow rates (60 and 90 L/h) and two dilution rates (0.1 and 0.22 h^{-1}) chosen to be below and above the range where sustained oscillations occur. The complete data set consists of the six measurement at four different steady-state operating conditions.

A constrained least-squares optimization problem is formulated to generate the parameter estimates. The

Table 6. Prediction Results for Two Parameters Estimated from Steady-State Data

	$D = 0.1$, $F = 90$		$D = 0.1$, $F = 60$		$D = 0.22$, $F = 90$		$D = 0.22$, $F = 60$	
measurement	plant	model	plant	model	plant	model	plant	model
dissolved O_2 [%]	72.5	63.84	71.02	75.73	42.15	48.79	39.95	44.98
evolved CO_2 [%]	0.78	0.88	1.11	0.95	1.52	1.36	2.18	2.0
evolved O_2 [%]	20	20	19	19.01	17.88	17.9	17.05	17.07
ethanol [g/L]	0.008	0.017	0.008	0.017	0.161	0.0937	0.172	0.116
glucose [g/L]	0.23	0.24	0.23	0.23	0.8	0.8	0.82	0.82
cell concentration [10^{10}/L]	2061	2040	2056	2030	1390	1380	1365	1360
SSE				2.745				

Table 7. Prediction Results for Three Parameters Estimated from Steady-State Data

	$D = 0.1$, $F = 90$		$D = 0.1$, $F = 60$		$D = 0.22$, $F = 90$		$D = 0.22$, $F = 60$	
measurement	plant	model	plant	model	plant	model	plant	model
dissolved O_2 [%]	72.5	74.8	71.02	73.45	42.15	41.67	39.95	39.47
evolved CO_2 [%]	0.78	0.73	1.11	1.03	1.52	1.54	2.18	2.2
evolved O_2 [%]	20.15	20.11	20.03	20.01	19.87	19.9	19.35	19.07
ethanol [g/L]	0.008	0.004	0.008	0.004	0.161	0.17	0.172	0.182
glucose [g/L]	0.23	0.23	0.23	0.23	0.8	0.8	0.82	0.82
cell concentration [10^{10}/L]	2061	2040	2056	2030	1390	1380	1365	1360
SSE				0.54				

Table 8. Initial and Final Parameters θ_1 for Four Parameters Estimated from Steady-State Data

		first set		second set	
parameter	plant values	initial	final	initial	final
m_{do}	10.25	10.5	10.68	10.75	10.66
m_{to}	4.55	4.7	4.58	4.3	4.35
m_e	1.54	1.5	1.63	1.7	1.65
γ_e	1.25	1.4	1.31	1.1	1.21

objective function is written as

$$f^s = \sum_{i=1}^{n} \sum_{j=1}^{m} w_{ij}^s (\hat{\bar{y}}_{ij} - \bar{y}_{ij})^2 \tag{29}$$

where f^s is the objective function for steady-state optimization; $n = 4$ is the number of data points; $m = 6$ is the number of measurements; w_{ij}^s is the weighting factor for the j-th measurement in the i-th data set; $\hat{\bar{y}}_{ij}$ is the steady-state prediction of the j-th measurement in the i-th data set; and $\bar{y}_{ij}$ is the corresponding plant measurement. The steady-state plant values are used as scaling factors:

$$w_{ij}^s = 1/\bar{y}_{ij}^2 \tag{30}$$

The decision variables in the optimization problem are the parameters θ_1 to be estimated. While only the measured variables are matched to plant data, the other state variables are constrained by the model. Therefore, the model equations form the equality constraints of the optimization problem:

$$\dot{\hat{x}} = f(\hat{x}, u, \theta_{fix}, \theta_1, \theta_2) \tag{31}$$

$$\hat{y} = h(\hat{x}, u, \theta_{fix}, \theta_1, \theta_2) \tag{32}$$

Inequality constraints arise from nonnegativity restrictions on the estimated parameters ($\theta_1 \geq 0$). The resulting nonlinear programming problem is solved using the routine fmincon in MATLAB.

Initial and optimized values for 1−3 estimated parameters are listed in Table 5. Although not shown here for the sake of brevity, a single estimated parameter (m_{do}) is insufficient to match the plant outputs. As shown in Table 6, significant prediction errors also are observed when a second parameter (m_{to}) is estimated. The most significant error occurs in the ethanol concentration. The prediction errors result in a relatively large normalized squared sum of errors (SSE) between the plant and model outputs. The parameter ranked third (m_e) has a strong effect on the predicted ethanol concentration. Consequently, three estimated parameters significantly improve the model predictions as reflected by the comparatively low SSE value in Table 7. The predicted outputs are sufficiently close to the plant outputs that the two sets would be almost indistinguishable by measurement.

Because the nonlinear optimization problem is nonconvex and has local minima, there is a range of initial values that produce a unique set of parameter estimates. Increasing the number of estimated parameters from two to three reduces this range roughly by 50%. The introduction of a fourth estimated parameter (γ_e) shrinks the range even further. As shown in Table 8, different values of the estimated parameters are obtained even with marginally different initial guesses. This occurs because γ_e is highly collinear with m_{to} and there is insufficient data to resolve their individual effects. Even though the SSE is reduced when the first set of optimized parameters in Table 8 is used to generate predictions (see Table 9), the nonuniqueness problem precludes the use of four

Table 9. Prediction Results for Four Parameters Estimated from Steady-State Data

	$D = 0.1$, $F = 90$		$D = 0.1$, $F = 60$		$D = 0.22$, $F = 90$		$D = 0.22$, $F = 60$	
measurement	plant	model	plant	model	plant	model	plant	model
dissolved O_2 [%]	72.5	72.91	71.02	72.02	42.15	44.69	39.95	41.51
evolved CO_2 [%]	0.78	0.78	1.11	1.13	1.52	1.48	2.18	2.03
evolved O_2 [%]	20.15	20.1	20.03	20.04	19.87	19.89	19.35	19.08
ethanol [g/L]	0.008	0.007	0.008	0.007	0.161	0.134	0.172	0.153
glucose [g/L]	0.23	0.22	0.22	0.22	0.8	0.8	0.82	0.82
cell concentration [10^{10}/L]	2061	2060	2056	2059	1390	1380	1365	1360
SSE				0.11				

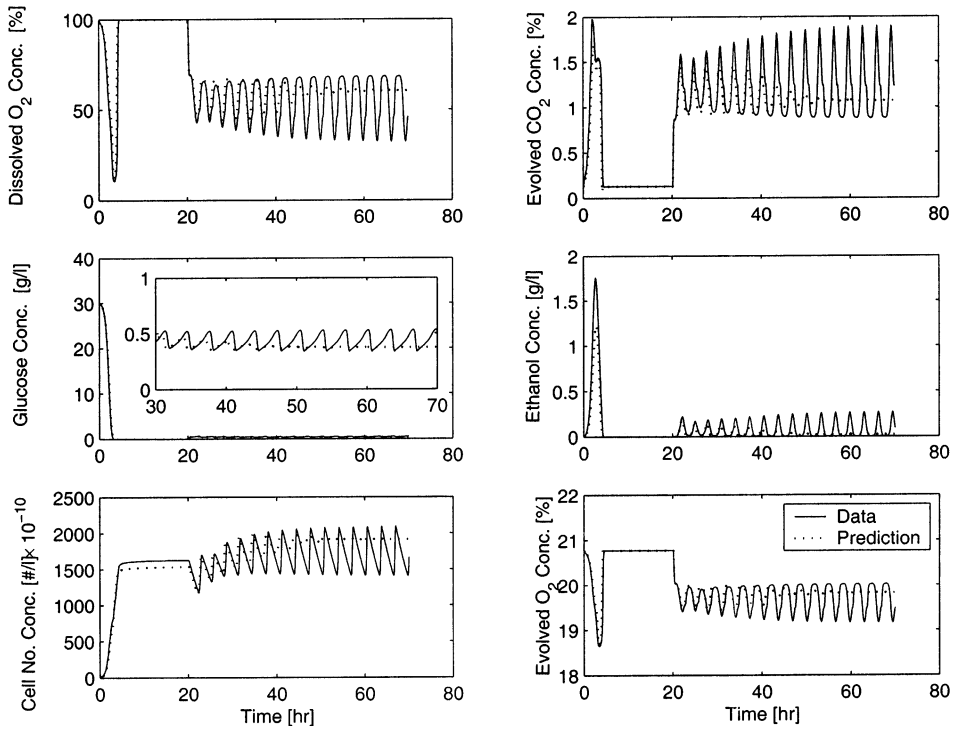

Figure 1. Simulation results for three parameters estimated from steady-state data (parameter values in Tables 4 and 5).

Table 10. Plant and Perturbed Parameters θ_2 for Dynamic Analysis

		parameters		
parameter	plant	1	2	3
m_{do}	10.25			
m_e	1.54	2.54		
K_t	0.01	0.021	0.021	
K_d	3.83	3.95	3.95	3.95
m_{to}	4.55	5.25	5.25	5.25
ϵ	7	10.94	10.94	10.94
m_o	1	0.65	0.65	0.65
γ	400	136.78	136.78	136.78
A	10	6.77	6.77	6.77
S_l	0.1	0.2	0.2	0.2
S_h	2	2.3	2.3	2.3
γ_e	1.25	1.55	1.55	1.55
ϵ_e	15	20.98	20.98	20.98
α_g	20	27.23	27.23	27.23
α_e	20	15.11	15.11	15.11

estimated parameters. In this case, the structure of the model allows only three parameters to be estimated from the available steady-state data.

Next the parameter estimates derived from steady-state data are evaluated for their ability to predict dynamic behavior. The simulated dynamic data used for analysis represents batch startup followed by a switch to continuous operation. Because sustained oscillation are observed in the plant following the switch, this data set tests the ability of the model parameters to support periodic solutions. The values of the parameters θ_2 and θ_1 used are listed in Tables 4 and 5, respectively. As shown in Figure 1, the model predictions are surprisingly good for the batch growth phase given that the dynamic model is not designed for this purpose. However, the model does not predict sustained oscillations following the switch to continuous operation at $t = 20$ h. This motivates the need to select and estimate model parameters based on dynamic considerations.

3.3. Parameter Estimation from Simulated Dynamic Data. Dynamic parameter estimation requires a data set that is rich in transient information including sustained oscillations. The optimization strategy used to solve the dynamic parameter estimation problem imposes computational restrictions on the data set. More specifically, the parameter estimates are obtained by iterating between a constrained nonlinear optimizer that minimizes the least-squares difference between the plant and model outputs over the dynamic data set and an integrator that computes dynamic model predictions for the current set of parameter estimates. Computational efficiency requires an information-rich data set that is collected over a relatively small time window. We have found that batch operation followed by a switch to continuous operation satisfies these requirements because the resulting data set contains information on the transient approach to sustained oscillations as well as the fully developed oscillations.

Table 11. Initial and Final Parameters θ_1 Estimated from Simulated Dynamic Data

	parameters					
	1		2		3	
parameter	initial	final	initial	final	initial	final
m_{do}	11	12.1688	11	11.75	11	10.813
m_e			1.8	1.6	1.8	1.52
K_t					0.02	0.585

On-line measurements of the dissolved oxygen concentration and the evolved oxygen and carbon dioxide concentrations are assumed to be available every minute. Off-line measurements of the glucose, ethanol, and cell number concentrations are assumed to be available every 6 min. Therefore, the dynamic data set is generated by sampling the plant outputs at 6-min intervals. The model is sampled at the same frequency to obtain the output predictions. The least-squares objective function is formulated as

$$f^d = \sum_{i=1}^{m} w_i^d \sum_{j=1}^{N} w_j^d [\hat{y}_i(j\Delta t) - y_i(j\Delta t)]^2 \qquad (33)$$

where f^d is the objective function for dynamic optimiza-

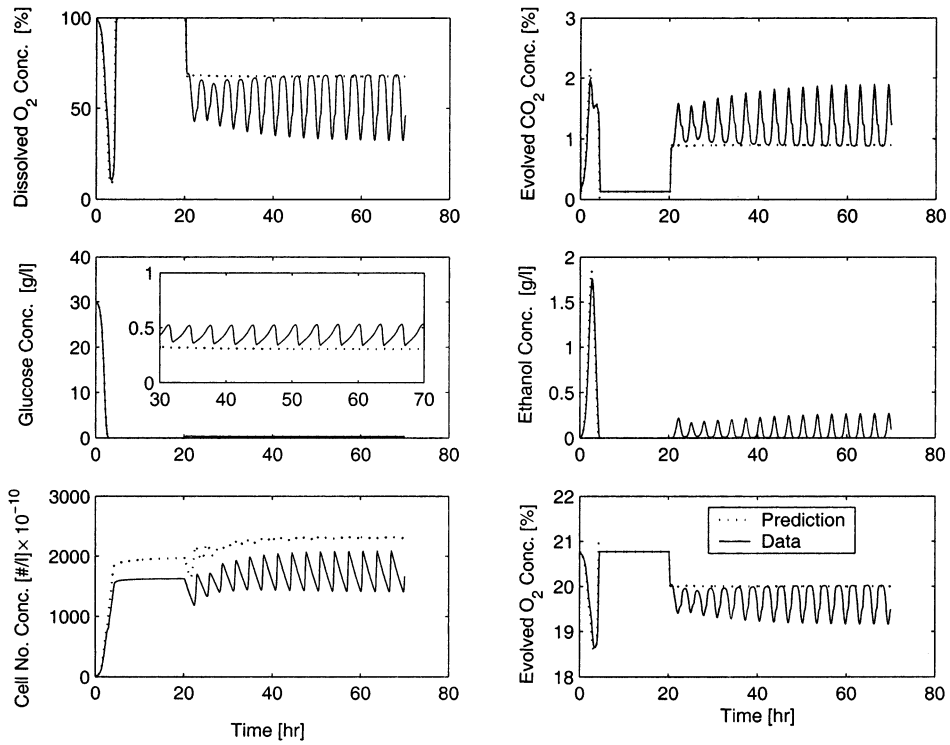

Figure 2. Simulation results for nominal parameter values in Table 10.

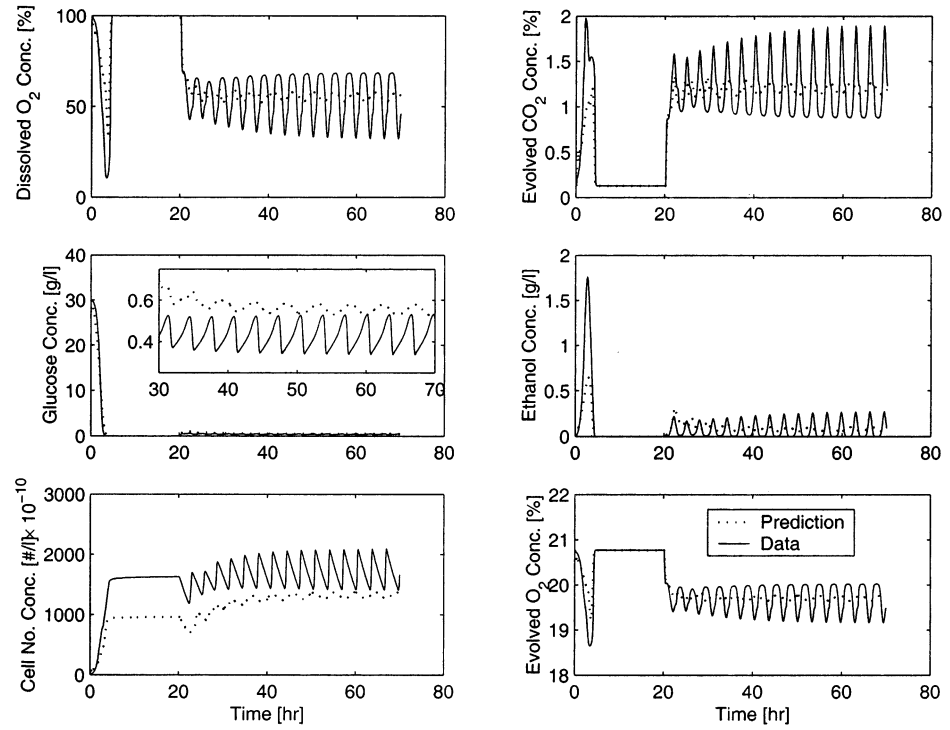

Figure 3. Simulation results for single parameter estimated from simulated dynamic data (parameter values in Tables 10 and 11).

tion; m is the number of measurements; N is the number of samples; Δt is the sampling time; $\hat{y}_i(j\Delta t)$ is the model prediction of the i-th variable at the j-th sampling time; $y_i(j\Delta t)$ is the corresponding plant value; w_i^d is a scaling weight the i-th variable chosen as in the steady-state case; and w_j^d is a dynamic weight that varies with the sample number. The dynamic weights are chosen as $w_j^d = j^2$ to place greater emphasis on matching the fully developed oscillations during continuous operation at the expense of the batch growth predictions. The Matlab functions fmincon and ODE15s are used for constrained optimization and integration, respectively. Iterations between the two subproblems are continued until the successive difference in the objective function is below a tolerance of 10^{-4}.

Values of the unknown model parameters θ_2 are generated from the plant values as in the steady-state case (see Table 10). The need for parameter estimation is demonstrated by simulating the model with the initial parameter values where $m_{do} = 11 \times 10^{-13}$ g and the remaining values are listed in the third column of Table 10. As shown in Figure 2, these values do not support

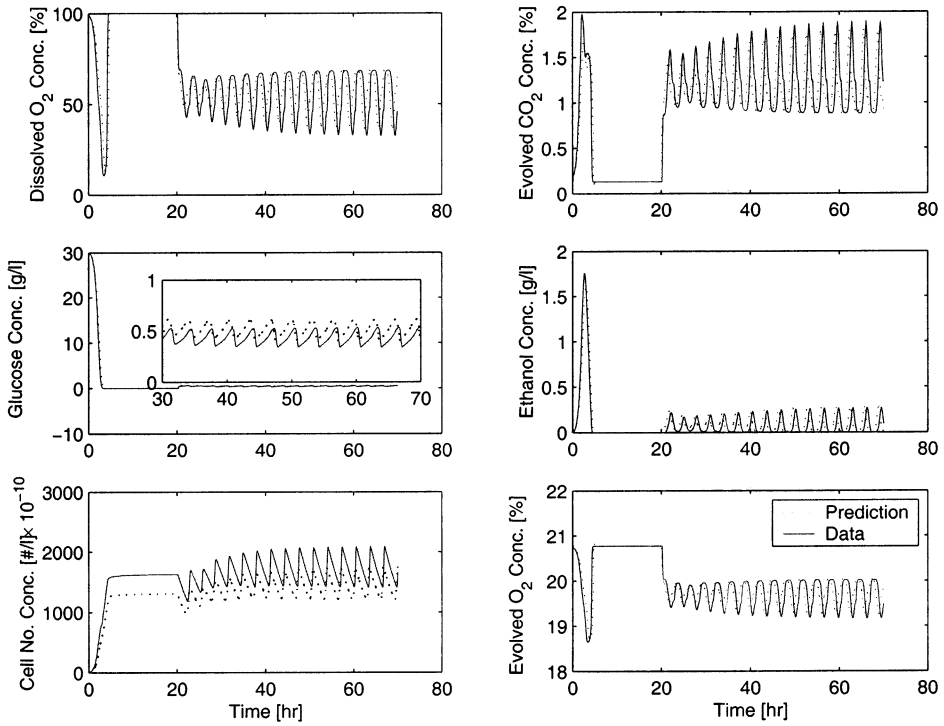

Figure 4. Simulation results for two parameters estimated from simulated dynamic data (parameter values in Tables 10 and 11).

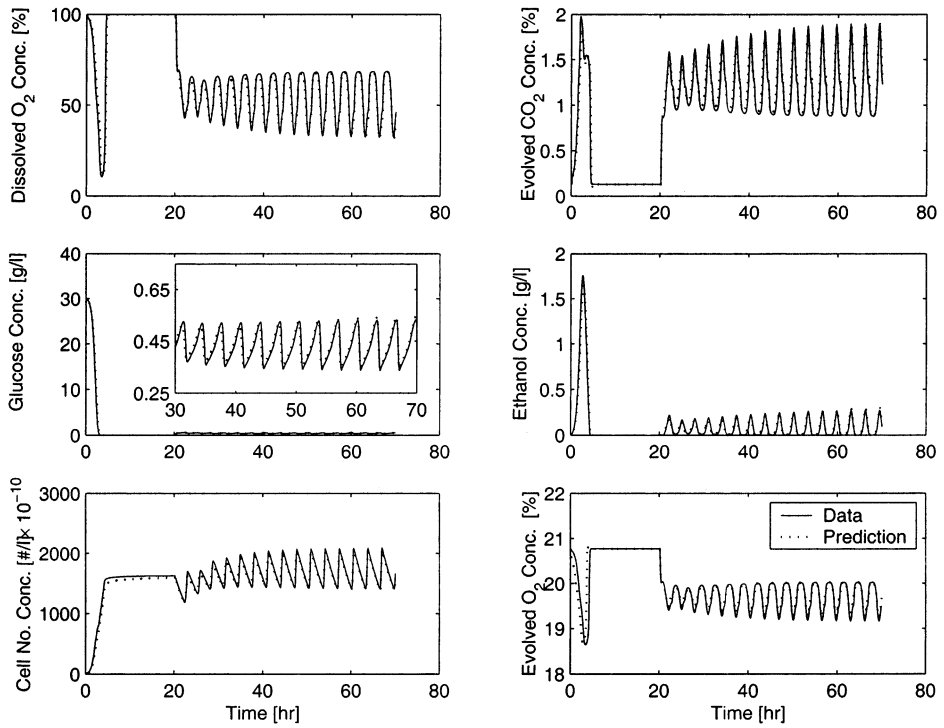

Figure 5. Simulation results for three parameters estimated from simulated dynamic data (parameter values in Tables 10 and 11).

Table 12. Steady-State Prediction Results for Three Parameters Estimated from Simulated Dynamic Data

	$D = 0.1$, $F = 90$		$D = 0.1$, $F = 60$		$D = 0.22$, $F = 90$		$D = 0.22$, $F = 60$	
measurement	plant	model	plant	model	plant	model	plant	model
dissolved O_2 [%]	72.5	71.44	71.02	69.9	42.15	42.18	39.95	40.01
evolved CO_2 [%]	0.78	0.81	1.11	1.15	1.52	1.52	2.18	2.18
evolved O_2 [%]	20.15	20.03	20.03	20.07	19.87	19.91	19.35	19.12
ethanol [g/L]	0.008	0.01	0.008	0.011	0.161	0.167	0.172	0.178
glucose [g/L]	0.23	0.23	0.22	0.23	0.8	0.81	0.82	0.83
cell concentration [10^{10}/L]	2061	2060	2056	2059	1390	1380	1365	1360
SSE				0.12				

sustained oscillations following the switch to continuous operation. Steady-state predictions of the measured outputs also are significantly different than the mean plant values.

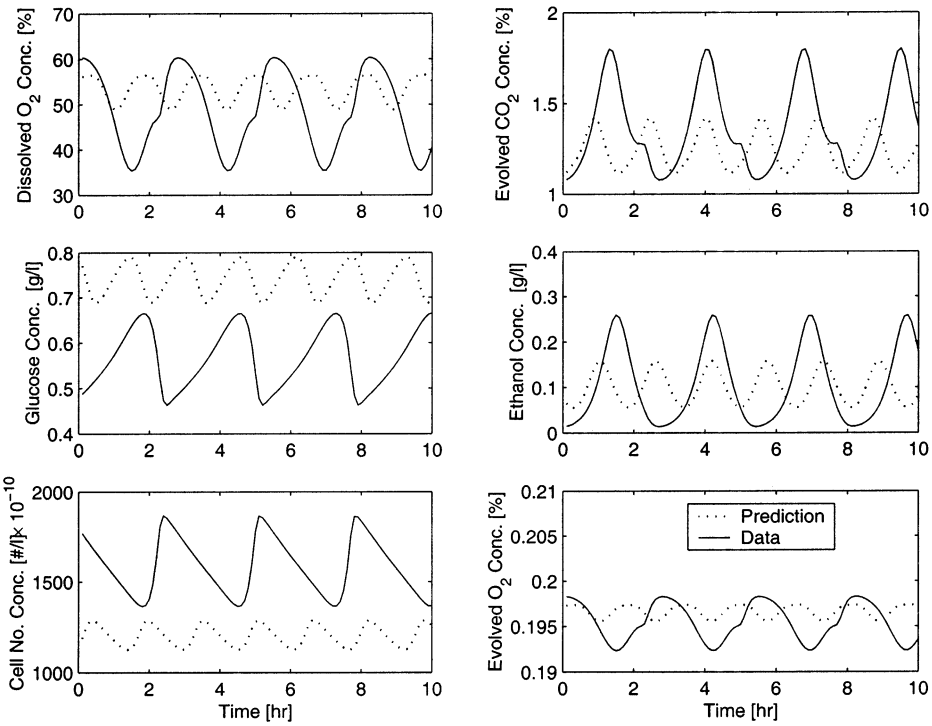

Figure 6. Comparison of plant and model oscillations at $D = 0.18$ h^{-1} using one parameter estimated from simulated dynamic data (parameter values in Tables 10 and 11).

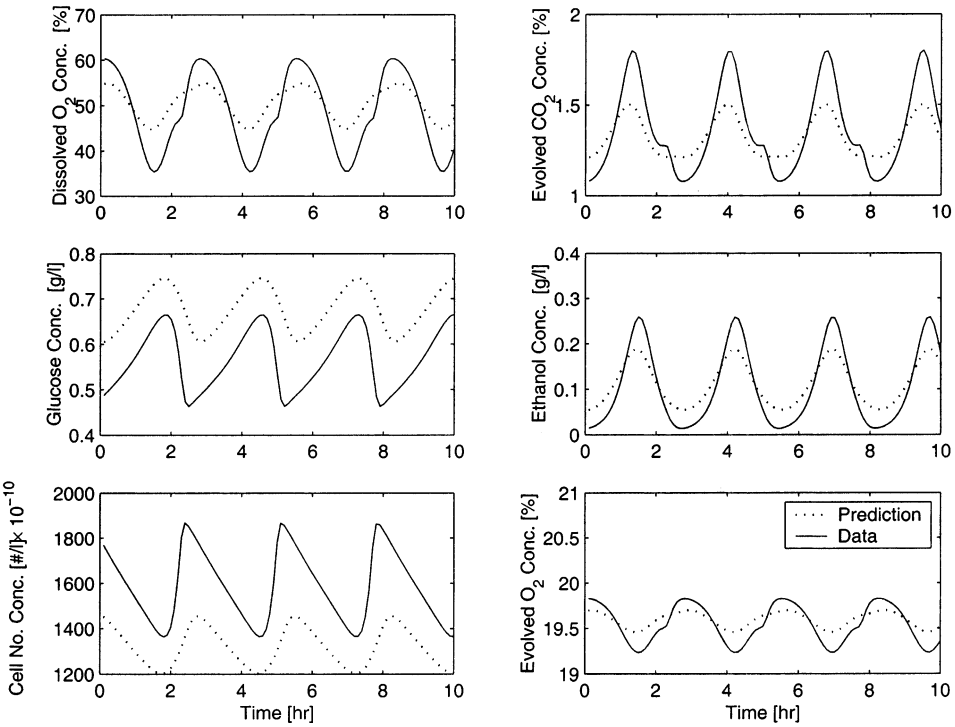

Figure 7. Comparison of plant and model oscillations at $D = 0.18$ h^{-1} using two parameters estimated from simulated dynamic data (parameter values in Tables 10 and 11).

Initial and optimized values for 1−3 estimated parameters are shown in Table 11. Figure 3 shows that estimation of a single parameter (m_{do}) does not produce sustained oscillations. However, the steady-state predictions are closer to the mean plant values than is achieved without estimation (see Figure 2). As shown in Figure 4, two estimated parameters (m_{do}, m_e) produce sustained oscillations. The second parameter m_e has a strong effect on the amount of ethanol produced. Because the ethanol concentration affects both the dissolved oxygen and carbon dioxide concentrations, these two variables are predicted accurately. By contrast, the glucose and cell number concentration predictions are poor because they are largely independent of the two estimated parameters. Figure 5 shows that three estimated parameters (m_{do}, m_e, K_t) produce a very good match of the plant data. This improvement is attributable to the inclusion of a third parameter K_t, which determines the amount of substrate consumed and therefore has a strong effect on the total cell number concentration.

Next the parameters estimated from dynamic data are evaluated for their ability to predict steady-state behav-

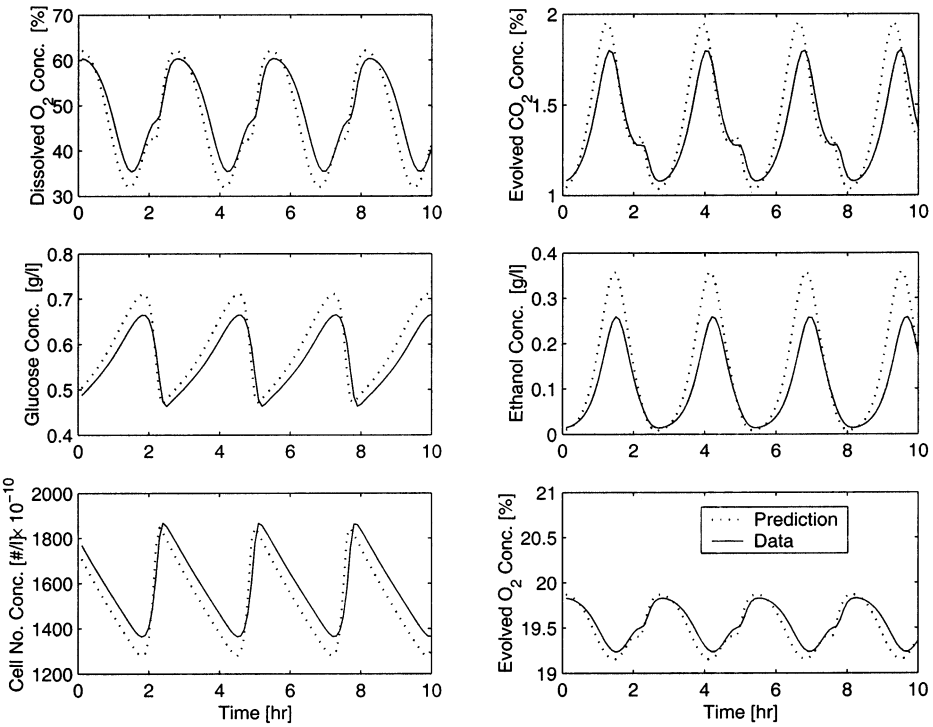

Figure 8. Comparison of plant and model oscillations at $D = 0.18\ h^{-1}$ using three parameters estimated from simulated dynamic data (parameter values in Tables 10 and 11).

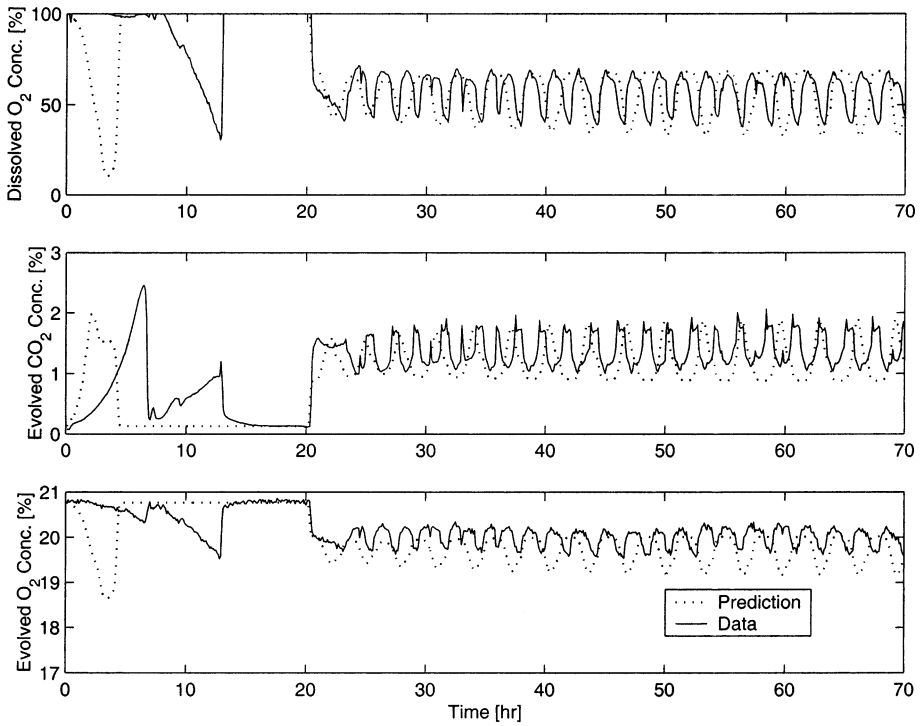

Figure 9. Comparison of simulated and experimental responses using nominal model parameter values in Tables 1 and 2.

ior. Table 12 shows that three parameters (m_{do}, m_e, K_t) estimated from reactor startup data predict the steady-state plant outputs more accurately than three parameters (m_{do}, m_{to}, m_e) estimated from steady-state data (see Table 7) as reflected by their relative SSE values. Apparently the dynamic data for batch and continuous operation provides more useful information about plant behavior than the steady-state data alone.

Thus far, prediction accuracy has been evaluated only for the training data set used in parameter estimation. To access extensibility of the model parameters, it is necessary to generate predictions for a different validation data set. For this purpose, the estimated parameters are evaluated for their ability to predict sustained

Table 13. Initial and Final Parameters θ_1 Estimated from Experimental Data

	parameters			
	3		4	
parameter	initial	final	initial	final
m_{do}	10.25	11.43	10.25	12.45
m_e	1.54	0.62	1.54	0.8
m_{to}	4.55	4.95	4.55	7.15
γ_e			1.25	1.81

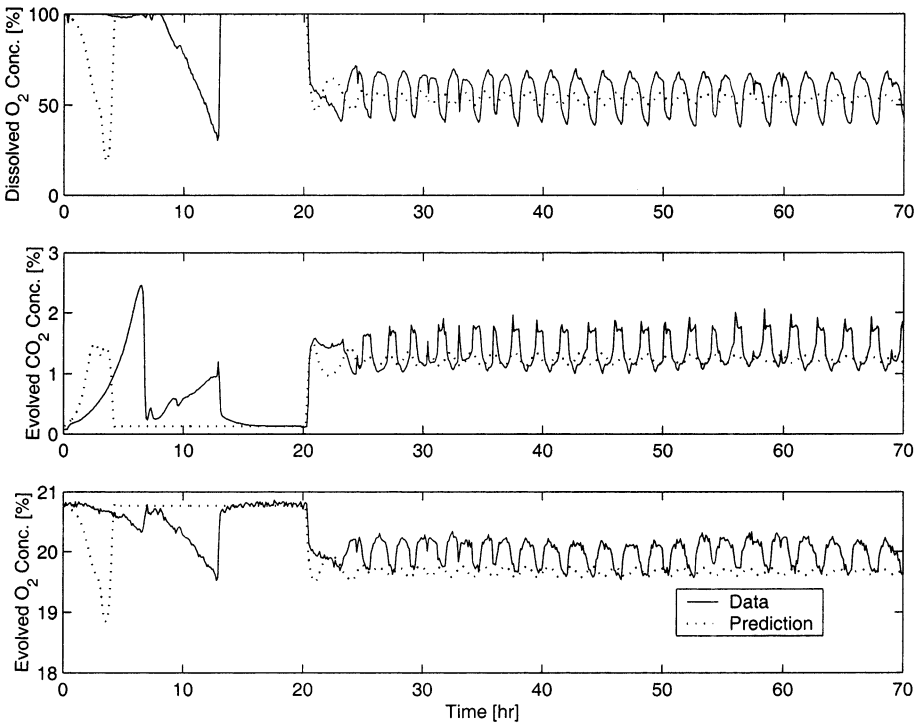

Figure 10. Comparison of simulated and experimental responses using three parameters estimated from experimental data (parameter values in Tables 10 and 13).

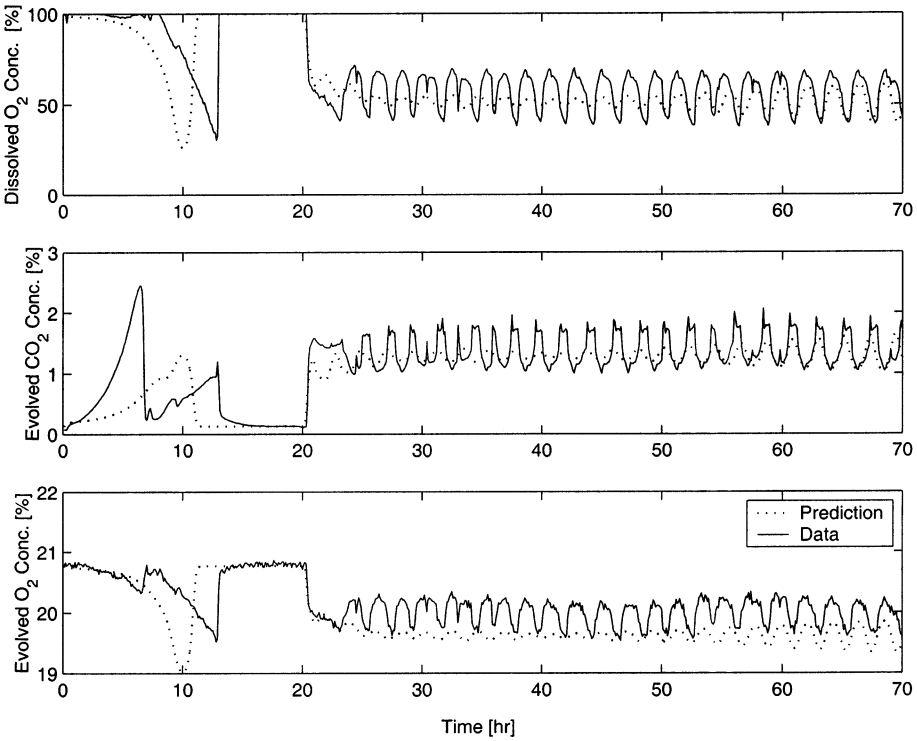

Figure 11. Comparison of simulated and experimental responses using four parameters estimated from experimental data (parameter values in Tables 10 and 13).

oscillations at $D = 0.18\ h^{-1}$ rather than $D = 0.15\ h^{-1}$ used in the training data set. As shown in Figure 6, a single estimated parameter (m_{do}) leads to sustained oscillations with significantly different mean values and amplitudes than observed in the plant. Figure 7 shows that two estimated parameters (m_{do}, m_e) yield improved predictions with respect to the evolved carbon dioxide, evolved oxygen, dissolved oxygen and ethanol concentrations. As observed with the training data set, the glucose and cell number concentrations are not well matched. Figure 8 shows that three estimated parameters (m_{do}, m_e, K_t) yield predictions that are in good agreement with the plant.

3.4. Estimation from Experimental Dynamic Data. The final and most important test of the parameter estimation strategy is the ability to match actual data collected in our laboratory. Figure 9 shows a comparison between model predictions generated using the nominal parameter values in Tables 1 and 2 and experimental data for a switch from batch to continuous operation. Although the amplitudes and mean values of the pre-

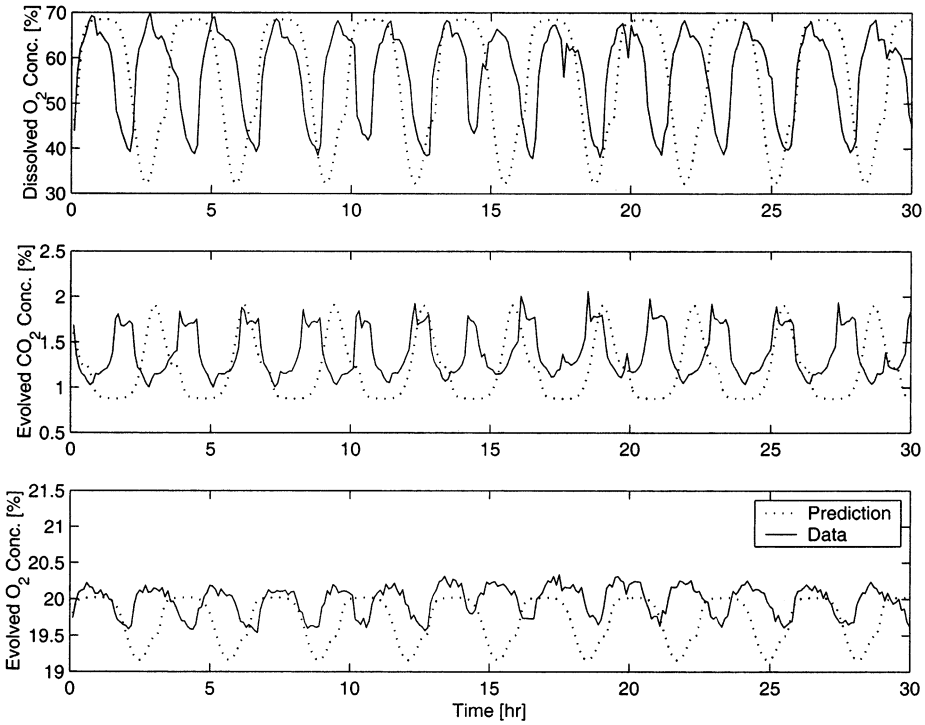

Figure 12. Comparison of simulated and experimental oscillations for nominal parameter values in Tables 1 and 2.

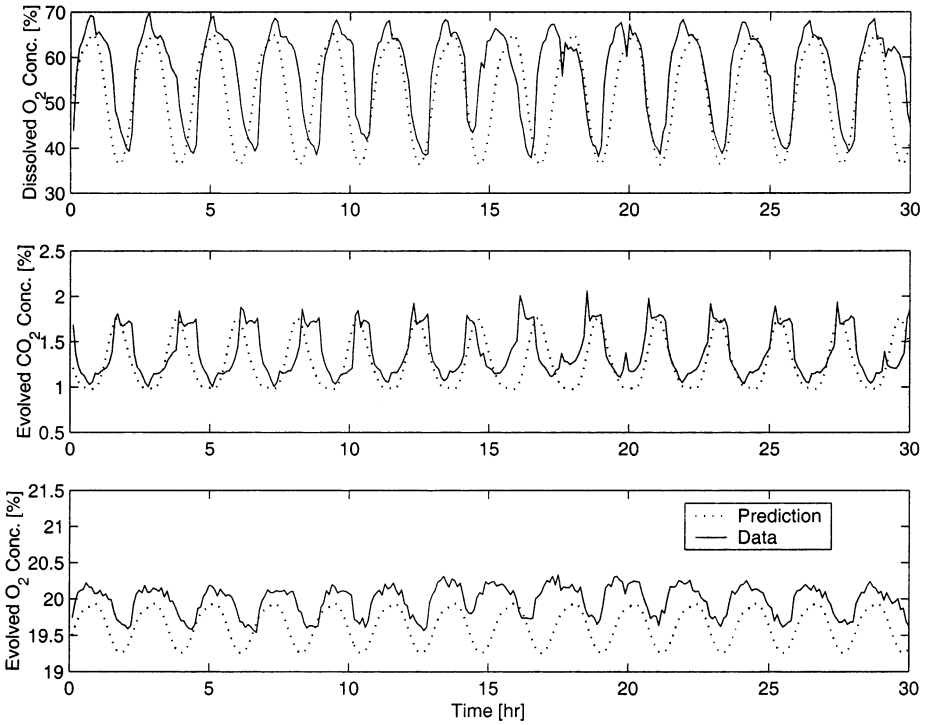

Figure 13. Comparison of simulated and experimental oscillations for four parameters estimated from experimental data (parameter values in Tables 10 and 13).

dicted variables are reasonably close to the experimental values, the predicted period is significantly too large. This motivates the need for parameter estimation to reduce the plant-model mismatch.

Formulation of the constrained nonlinear optimization problem is identical to that presented earlier for simulated dynamic data except that only three measured outputs (dissolved oxygen, evolved carbon dioxide, and evolved oxygen concentrations) are available. The previous simulation tests suggest that at least three parameters must be estimated to capture oscillatory dynamics. In this case, the parameter rankings for experimental dynamic data (see Table 3) are used to select the estimated parameters θ_1. The parameters θ_2 are fixed at the values listed in the second column of Table 10. Initial and optimized values for three and four estimated parameters are listed in Table 13.

As shown in Figure 10, three estimated parameters (m_{do}, m_e, m_{to}) do not produce sustained oscillations following the switch from batch to continuous operation. Furthermore, the steady-state prediction of the evolved oxygen concentration is significantly different from the

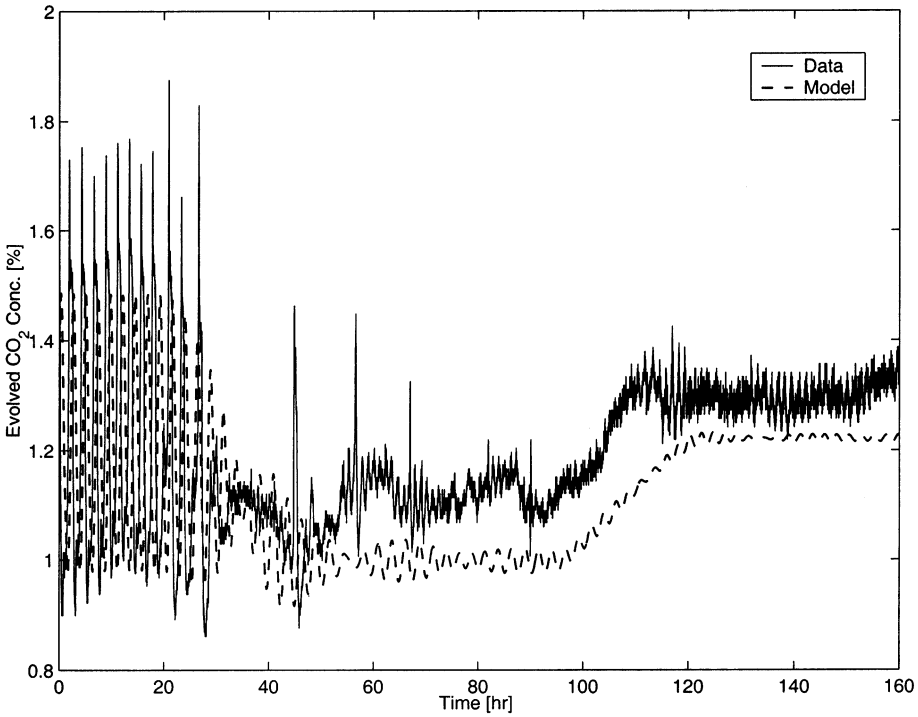

Figure 14. Comparison of simulated and experimental dilution rate ramp tests using four parameters estimated from experimental data (parameter values in Tables 10 and 13).

mean experimental value. The inability to satisfactorily predict the experimental data using three estimated parameters may be attributable to structural limitations of the model. To address this problem, the parameter γ_e ranked fourth is included in the set θ_1 of estimated parameters. While some improvement in the batch growth phase are evident in Figure 11, the model dynamics are considerably slower than the dynamics of the real system. As a result, the predicted oscillations take much longer to become fully developed. To evaluate the ability of the model to capture fully developed oscillations, model predictions generated with the nominal parameter values and with the four estimated parameter values are compared with experimental data in Figures 12 and 13, respectively. The predicted oscillations are initiated to be in phase with the experimental oscillations. The results demonstrate that estimation of four parameters yields significant improvements in the model predictions.

The transient response of the evolved carbon dioxide concentration for a series of dilution rate changes is used to evaluate extensibility of the estimated model parameters to other operating conditions. The dilution rate is ramped sufficiently slowly that the fermenter remains in a quasi-steady state (*25*). Experimental data is generated by setting $D = 0.13$ h^{-1} for the first 20 h, followed by a 24-h ramp down to $D = 0.1$ h^{-1}, operation at $D = 0.1$ h^{-1} between 44 and 96 h, followed by a 24-h ramp up to $D = 0.13$ h^{-1} and subsequent operation at $D = 0.13$ h^{-1} until $t = 160$ h. The experimental data shown in Figure 14 suggests the existence of both stable stationary and oscillatory states at $D = 0.13$ h^{-1}. The four estimated parameters do not support sustained oscillations at this dilution rate. However, sustained oscillations are predicted at $D = 0.134$ h^{-1}, which is very close to the experimental value. Therefore, dynamic model simulation is performed using $D = 0.134$ h^{-1} as the upper value of the dilution rate; otherwise the dilution rate sequence is identical to that used experimentally. Figure 14 shows that the model captures the damping behavior at lower dilution rates and the very small decaying oscillations when the dilution rate is returned back to the oscillatory range. These results demonstrate that the proposed model combined with parameter estimation can satisfactorily predict the dynamic behavior of *S. cerevisiae*.

4. Summary and Conclusions

A cell population model with a simple structured description of the extracellular environment has been presented for the microorganism *S. cerevisiae*. As compared to completely unstructured models, the major advantage of the proposed model is that it contains extracellular variables that can be compared directly to readily available laboratory measurements. This facilitates the development of model parameter estimation techniques that allow quantitative prediction of experimental data. A parameter selection method previously developed in our group is shown to be useful for identifying subsets of parameters that can be estimated reliably from steady-state measurements. The method does not directly determine the number of parameters that can be uniquely determined from a given data set. Tests with simulated steady-state data reveal that three model parameters can be estimated reliably, while the inclusion of a fourth parameter leads to collinearity and convergence problems. Parameter estimation from steady-state data is shown to be inadequate for predicting oscillatory yeast dynamics.

The parameter selection method has been extended to fully developed oscillations where model behavior is characterized by the mean, amplitude and period of each measured variable. Parameter estimation with simulated dynamic data improves the prediction of oscillatory dynamics as compared to the steady-state case. Three parameters are shown to be sufficient to capture the plant behavior. Application of the proposed methodology to experimental data collected in our laboratory leads to significant improvements in the prediction of dynamic behavior including fully developed oscillations. This work

demonstrates the utility of simple segregated model for describing the essential features of yeast culture dynamics.

Notation

C dissolved carbon dioxide concn [g/L]
C^* saturation carbon dioxide concn [g/L]
C_{in} carbon dioxide partial pressure in air feed stream [atm]
C_{out} carbon dioxide partial pressure in gas exhaust stream [atm]
D dilution rate [h^{-1}]
E ethanol concn [g/L]
E' intracellular ethanol concn [g/L]
E_f ethanol feed concn [g/L]
F volumetric flow rate of air [L/h]
G glucose concn [g/L]
G' intracellular glucose concn [g/L]
G_f glucose feed concn [g/L]
H_o Henry's law constant for oxygen [g/L/atm]
H_c Henry's law constant for carbon dioxide [g/L/atm]
K_d slope of division mass with respect to substrate concn in saturation function [L]
K_{eo} ethanol oxidation rate [g/h]
K_{gf} glucose fermentation rate [g/h]
K_{go} glucose oxidation rate [g/h]
K_{inhib} inhibition constant for ethanol oxidation [g/L]
k_{lc} carbon dioxide mass transfer coefficient [m/h]
K_{lo} oxygen mass transfer coefficient [m/h]
K_{meo} ethanol saturation constant for ethanol oxidation [g/L]
K_{mgf} glucose saturation constant for glucose fermentation [g/L]
K_{mgo} glucose saturation constant for glucose oxidation [g/L]
K overall single cell growth rate [g/h]
K_t slope of transition mass with respect to substrate concn in saturation function [L]
O dissolved oxygen concn [g/L]
O^* saturation dissolved oxygen concn [g/L]
O_{out} oxygen partial pressure in gas exhaust stream [atm]
O_{in} oxygen partial pressure in the feed stream [atm]
R gas constant [L atm/mol/K]
S_l, S_h substrate concn limits in the saturation functions [g/L]
S' intracellular substrate concn [g/L]
T absolute temperature [K]
V_g reactor gas-phase volume [L]
V_l reactor liquid-phase volume [L]
W distribution of cell mass [no./g]
$\hat{x}$ model state vector
Y_{eo} yield coefficient for ethanol oxidation [g/g]
Y_{gf} yield coefficient for glucose fermentation [g/g]
Y_{go} yield coefficient for glucose oxidation [g/g]
a interfacial area per unit liquid volume [m^{-1}]
f ethanol production function
f^d objective function for parameter estimation from dynamic data
f^s objective function for parameter estimation from steady-state data
k_{go} dissolved oxygen saturation constant for glucose oxidation [g/L]
k_{eo} dissolved oxygen saturation constant for ethanol oxidation [g/L]
m_e mass cells must attain to produce ethanol [g]
m_d^* cell division mass [g]
m_t^* cell transition mass [g]
m_o mass above m_t^* a mother cell must gain before division is possible [g]
m_{do} saturation value for division mass [g]
m_{to} saturation value for transition mass [g]
p newborn cell mass distribution function
u input vector
y measured variables in plant
$\hat{y}$ measured variables as predicted by model
α_e filter constant for ethanol [h^{-1}]
α_g filter constant for glucose [h^{-1}]
Γ division intensity function
γ preexponential factor in division intensity function
γ_e preexponential factor in ethanol production function
ϵ division intensity function slope
ϵ_e ethanol production function slope
θ_1 unknown parameter values that are estimated
θ_2 unknown parameter values that are not estimated
θ_{fix} parameter values derived from literature
θ_{est} parameter values available for estimation
λ eigenvalues of the principle component matrix
μ_{meo} maximum ethanol oxidation rate [g/h]
μ_{mgo} maximum glucose oxidation rate [g/h]
μ_{mgf} maximum glucose fermentation rate [g/h]

Acknowledgment

Partial financial support from the LSU Chemical Engineering Department though a Gordon and Mary A. Cain Graduate Student Fellowship is gratefully acknowledged.

References and Notes

(1) Alberghina, L.; Ranzi, B. M.; Porro, D.; Martegani, E. Flow cytometry and cell cycle kinetics in continuous and fed-batch fermentations of budding yeast. *Biotechnol. Prog.* **1991**, *7*, 299−304.
(2) Beuse, M.; Bartling, R.; Kopmann, A.; Diekmann, H.; Thoma, M. Effect of dilution rate on the mode of oscillation in continuous culture of *Saccharomyces cerevisiae*. *J. Biotechnol.* **1998**. *61*, 15−31.
(3) Chen, C.-I.; McDonald, K. A. Oscillatory behavior of *Saccharomyces cerevisiae* in con- tinuous culture: II. Analysis of cell synchronization and metabolism. *Biotechnol. Bioeng.* **1990**, *36*, 28−38.
(4) Chen, C.-I. McDonald, K. A.; Bisson, L. Oscillatory behavior of *Saccharomyces cerevisiae* in continuous culture: Effects of pH and nitrogen levels. *Biotechnol. Bioeng.* **1990**, *36*, 19−27.
(5) Martegani, E. Porro, D.; Ranzi, B. M.; Alberghina, L. Involvement of a cell size control mechanism in the induction and maintenance of oscillations in continuous cultures of budding yeast. *Biotechnol. Bioeng.* **1990**, *36*, 453−459.
(6) Munch, T.; Sonnleitner, B.; Fiechter, A. Cell cycle of synchronously cultivated *Saccharomyces cerevisiae*. In *Biochemical Engineering Stuttgart*; Reuss, M., Chmiel, H., Gilles, E., Knackmuss, H., Eds.; Gustav Fischer: Stuttgart, New York, 1991; pp 373−376.
(7) Munch, T.; Sonnleitner, B.; Fiechter, A. The decisive role of the *Saccharomyces cerevisiae* cell cycle behavior for dynamic growth characterization. *J. Biotechnol.* **1992**, *22*, 329−352.
(8) Munch, T.; Sonnleitner, B.; Fiechter, A. New insights into the synchronization mechanism with forced synchronous cultures of *Saccharomyces cerevisiae*. *J. Biotechnol.* **1992**, *24*, 299−313.

(9) Parulekar, S. J.; Semones, G. B.; Rolf, M. J.; Lievense, J. C.; Lim, H. C. Induction and elimination of oscillations in continuous cultures of *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **1986**, *28*, 700−710.
(10) Porro, D. E.; Martegani, B.; Ranzi, M.; Alberghina, L. Oscillations in continuous cultures of budding yeasts: A segregated parameter analysis. *Biotechnol. Bioeng.* **1988**, *32*, 411−417.
(11) Strassle, C.; Sonnleitner, B.; Fiechter, A. A predictive model for the spontaneous synchronization of *Saccharomyces cerevisiae* grown in continuous culture. I. Concept. *J. Biotechnol.* **1988**, *7*, 299−318.
(12) Strassle, C.; Sonnleitner, B.; Fiechter, A. A predictive model for the spontaneous syn- chronization of *Saccharomyces cerevisiae* grown in continuous culture. II. Experimental verification. *J. Biotechnol.* **1989**, *9*, 191−208.
(13) Keulers, M.; Satroutdinov, A. D.; Suzuki, T.; Kuriyama, H. Synchronization affector of autonomous short-period-sustained oscillation of *Saccharomyces cerevisiae*. *Yeast* **1996**, *12*, 673−682.
(14) Keulers, M.; Suzuki, T.; Satroutdinov, A. D.; Kuriyama, H. Autonomous metabolic oscillations in continuous culture of *Saccharomyces cerevisiae* grown on ethanol. *FEMS Microbiol. Lett.* **1996**, *142*, 253−258.
(15) Jones, K. D.; Kompala, D. S. Cybernetic modeling of the growth dynamics of *Saccharomyces cerevisiae* in batch and continuous cultures. *J. Biotechnol.* **1999**, *71*, 105−131.
(16) Hjortsø, M. A.; Nielsen, J. Population balance models of autonomous microbial oscillations. *J. Biotechnol.* **1995**, *42*, 255−269.
(17) Srienc, F.; Dien, B. S. Kinetics of the cell cycle of *Saccharomyces cerevisiae*. *Ann. N.Y. Acad, Sci.* **1992**, 59−71.
(18) Zhu, G.-Y.; Zamamiri, A. M.; Henson, M. A.; Hjortsø, M. A. Model predictive control of continuous yeast bioreactors using cell population models. *Chem. Eng. Sci.* **2000**, *55*, 6155−6167.
(19) Eakman, J. M.; Fredrickson, A. C.; Tsuchiya, H. H. Statistics and dynamics of microbial cell populations. *Chem. Eng. Prog. Symp. Ser.* **1996**, *62*, 37−49.
(20) Hjortsø, M. A. Periodic forcing of microbial cultures. A model for induction synchrony. *Biotechnol. Bioeng.* **1987**, *30*, 825−835.
(21) Woehrer, W.; Roehr, M. Regulatory aspects of baker's yeast metabolism in aerobic fed- batch cultures. *Biotechnol. Bioeng* **1981**, *23*.
(22) Frykman, S. *Single Cell Secretion Rates of Saccharomyces cerevisiae.* Ph.D. Thesis, University of Minnesota, 1999.
(23) Rizzi, M.; Theobald, U.; Querfurth, E.; Rohrhirsch, T.; Baltes, M.; Reuss, M. *In Vivo* investigations of glucose transport in *Saccharomyces cerevisiae*. *Biotechnol. Bioeng.* **1996**, *49*, 316−327.
(24) Von Meyenburg, H. K.; Feichter, A. Regulation phenomena of gaseous metabolism of aerobically growing baker's yeast. In *Proceedings of the 2nd International Symposium on Yeast*; Bratislava, Slovakia, 1968; pp 377−385.
(25) Zamamiri, A. M.; Birol, G.; Hjortso, M. A. Multiple stable states and hysteresis in continuous, oscillating cultures of budding yeast. *Biotechnol. Bioeng.* **2001**, *75*, 305−312.
(26) Jacquez, J. A.; Greif, P. Numerical parameter identifiablity and estimability: Integrating identifiablity, estimability, and optimal sampling design. *Math. Biosci* **1985**, *76*, 201−327.
(27) Mehra, R. N. Optimal inputs for linear system identification. *IEEE Trans. Auto. Cont.* **1974**, *19*, 192−200.
(28) Sandik, C. A.; McAuley, K. B.; McLellan, P. J. Selection of parameters for updating in on-line models. *Ind. Eng. Chem. Res.* **2001**, *40*, 3936−3950.
(29) Li, R.; Henson, M. A.; Kurtz, M. J. Selection of model parameters for off-line parameters estimation. *IEEE Trans. Control Syst. Technol.* Submitted for publication.
(30) Anderson, T. W. *An Introduction to Multivariate Statistical Analysis.* John Wiley & Sons: New York, 1958.
(31) Dunteman, G. H. *Principal Component Analysis*; Sage Publications Inc.: Newbury Park, CA, 1989.

Accepted for publication June 5, 2002.

BP020083I