Modeling in microbial batch culture and its parameter identification

Zhaohua Gong
Chongyang Liu
Enmin Feng, Dalian University of Technology

Available at: https://works.bepress.com/chongyang_liu/11/
Modeling in microbial batch culture and its parameter identification

Zhaohua Gong, Chongyang Liu and Enmin Feng

Abstract—In this paper, the nonlinear dynamical system of batch fermentation is investigated in the bioconversion of glycerol to 1,3-propanediol(1,3-PD) by Klebsiella pneumoniae. Taking account of the kinetic behavior and experimental results in the batch cultures, we propose a two-stage dynamical system to formulate the fermentation process. Then some properties of the proposed system are proved. In view of the big errors between observations and numerical simulation results, we subsequently establish a parameter identification model to identify parameters in the system. The identifiability of the model is also discussed. Finally, in order to find the optimal parameters of the identification model, an improved simulated annealing algorithm combined with Hook-Jeeves local search is constructed. Numerical results show that the two-stage system can describe the factual fermentation better and the optimization algorithm is valid.

I. INTRODUCTION

1,3-propanediol(1,3-PD) possesses potential applications on a large commercial scale, especially as a monomer of polyesters or polyurethanes, its microbial production is recently paid attention to for its low cost, high production and no pollution, etc. It is considered to be one of the bulk chemicals, which is likely to be produced by bioprocesses on large scales. These works include the quantitative description of the cell growth kinetics of multiple-inhibitions, the metabolic overflow kinetics of substrate consumption and product formation [1], [2], [3], feeding strategy of glycerol in fed-batch culture [4] and model analysis and simulations to determine the optimal operation conditions [5] etc. Moreover, the identification models and optimization algorithms are investigated with respect to batch, continuous, and fed-batch cultures in [6], [7], [8], respectively. The methods of glycerol bioconversion to 1,3-PD consist of batch, continuous and fed-batch cultures, among which the batch culture is fundamental. It’s known that the exact formulation of batch process is favorable to the study of continuous and fed-batch cultures. According to the cell growth kinetics of multiple inhibitions, the metabolic overflow kinetics of substrate consumption and product formation, the process of batch culture is composed of developmental phase, growth period and stationary phase. However, the previous nonlinear systems can just well formulate the first two phases and the errors between the experimental data and simulated values are very large in stabilization period. That is to say, the previous models can’t well describe the process of the batch culture.

Hence the purpose of this paper is to present a dynamical system of glycerol bioconversion to 1,3-PD, which can formulate the fermentation process better. Firstly, a two-stage dynamical system is proposed according to the kinetic behavior and experimental results in the batch culture. Subsequently we prove some prime properties of the above system, such as existence, uniqueness and boundedness of the solution. In view of the errors between the observations and computational values, we utilize a least squares technique and establish a parameter identification model to identify the parameters in the system. Moreover, we discuss the identifiability of the parameters. Finally, to find the optimal parameters of the identification model quickly, an improved simulated annealing algorithm combined with Hook-Jeeves local search is constructed. Numerical results show that the two-stage system is fit for the factual fermentation better and the optimization algorithm is valid.

The rest of this paper is organized as follows. In Section II, a two-stage nonlinear system of batch fermentation is presented. In Section III, some properties of the solution to the system are proved and the parameter identification model is established. In Section IV, we construct an improved simulated annealing algorithm to find the optimal parameters of the two-stage system. In addition, a numerical example is given. Finally, Section V concludes some conclusions and future works.

II. TWO-STAGE DYNAMICAL SYSTEM IN BATCH CULTURE

In batch culture, a quantity of biomass and glycerol are added to the reactor only once and stirred uniformly under given conditions. During the process of the culture, the concentration of the glycerol decreases gradually and tends to zero finally. According to the actual fermentation process, we make the following assumptions.

\( (H_1) \): Don’t input or output the biomass, substrate and products in reactor during the process of batch fermentation.

\( (H_2) \): The concentrations of reactants are uniform in reactor.

Under the above assumptions, the simplified mass balances of biomass, substrate and products in batch culture are written as the following nonlinear dynamical system [9]

\[
\begin{align*}
\dot{x}_1(t) &= \mu x_1, \\
\dot{x}_2(t) &= -q_2 x_1, \\
\dot{x}_i(t) &= q_i x_1, & i = 3, 4, 5.
\end{align*}
\]
Where \( t_f \in (0, +\infty) \) is the terminal moment of the batch culture. \( x_1(t), x_2(t), x_3(t), x_4(t), x_5(t) \) respectively denote the concentrations of biomass, glycerol, 1,3-PD, acetic acid and ethanol concentrations at \( t \) in reactor. The specific growth rate of cells \( \mu \), specific consumption rate of substrate \( q_i \) and specific formation rates of products \( q_i, i = 3, 4, 5 \) are expressed by (2)-(4), respectively.

\[
\begin{align*}
\mu &= \mu_\text{m} - \frac{x_2(t)}{x_2(t) + x_3(t)} + k_2 \sum_{i=2}^{5} \left( 1 - \frac{x_i(t)}{x_i^\text{m}} \right)^{n_i},
\end{align*}
\]

(2)

\[
q_2 = m_2 + \frac{\mu}{Y_2},
\]

(3)

\[
q_i = m_i + \frac{\mu}{Y_i}, \quad i = 3, 4, 5.
\]

(4)

The system (1) had been used by [6]. Numerical results in [6] show that the system (1) can only describe the developmental and vegetation phases well. However, in the stationary phase, the errors between the experimental data and computational values are very large. To formulate the process of batch fermentation better, we revise the model of the stationary phase according to the experimental data in the fermentation process. Let \( t_g \in [0, t_f] \) be the moment after which the system reaches the stationary phase. Then the time interval of culture process \( [0, t_f] \) is divided into two phases, i.e., \( [0, t_g) \) is the time interval of developmental and growth periods and \([t_g, t_f)\) is the one of stationary phase. So a two-stage nonlinear dynamical system is

\[
\begin{align*}
\dot{x}_1(t) &= \mu x_1, \\
\dot{x}_2(t) &= -q_2 x_1, \\
\dot{x}_3(t) &= q_1 x_1, \\
\dot{x}_4(t) &= \mu e^{-a_1(t-t_g)} x_1, \\
\dot{x}_5(t) &= -q_2 e^{-a_2(t-t_g)} x_1, \\
\dot{x}_j(t) &= q_3 e^{-a_3(t-t_g)} x_1, \\
x_j(0) &= x_{j0}, \quad x_j(t_g^+) = x_j(t_g), \quad j \in I_5.
\end{align*}
\]

(5)

Where \( t_g \) is given by the experimental results; \( x_j(t_g^+) \) denotes the right limit of concentration at time \( t_g \).

Under anaerobic conditions at 37°C and pH 7.0, the maximum specific growth rate \( \mu_\text{m} \) is 0.67h\(^{-1}\). The critical concentrations of biomass, glycerol, 1,3-PD, acetate and ethanol for cell growth are \( x_1^\text{m} = 10gL^{-1}, x_2^\text{m} = 2039mmolL^{-1}, x_3^\text{m} = 939.5mmolL^{-1}, x_4^\text{m} = 1026mmolL^{-1} \) and \( x_5^\text{m} = 360.9mmolL^{-1} \), respectively.

Let the parameters to be identified in (5) be

\[
u := \begin{pmatrix} k_2, m_2, m_3, m_4, m_5, Y_2, Y_3, Y_4, Y_5, n_2, n_3, n_4, n_5, a_1, a_2, a_3, a_4, a_5 \end{pmatrix} \in \mathbb{R}^{18}.
\]

The upper and lower bounds of the parameter \( \nu \) in consideration are respectively

\[
u^b = \begin{pmatrix} 100, 10, 10, 10, 10, 100, 100, 100, 5, 5, 5, 5, 3, 3, 3, 3 \end{pmatrix}.
\]

and

\[
u^a = \begin{pmatrix} 0.01, 0.01, -10, -10, -10, 0.0001, 1, 1, 1, 0, 0, 0, 0.0001, 0.01, 0, 0, 0 \end{pmatrix}.
\]

Hence, the admissible set of the parameters is defined as \( U_{\text{ad}} := \bigcap_{i=1}^{18} [\nu_i^a, \nu_i^b] \).

Let \( x(t) := (x_1(t), \cdots, x_5(t))^T \in \mathbb{R}^5, t \in [t_0, t_f] \), be the state vector and \( x(0) := (x_{01}, \cdots, x_{05})^T \in \mathbb{R}^5 \) be the initial state. Furthermore, let

\[
\begin{align*}
f^1(x(t), u) := \begin{pmatrix} \mu x_1(t) - q_2 x_1(t), q_3 x_1(t), q_4 x_1(t), q_5 x_1(t) \end{pmatrix}^T,
\end{align*}
\]

(6)

and

\[
\begin{align*}
f^2(x(t), u) := \begin{pmatrix} (\mu e^{-a_1(t-t_g)} x_1(t) - q_2 e^{-a_2(t-t_g)} x_1(t), q_3 e^{-a_3(t-t_g)} x_1(t), q_4 e^{-a_4(t-t_g)} x_1(t), q_5 e^{-a_5(t-t_g)} x_1(t) \end{pmatrix}^T.
\end{align*}
\]

(7)

Then, the two-stage system (5) of the batch culture can be rewritten as

\[
\begin{align*}
\begin{cases}
\dot{x}(t) &= f^1(x(t), u), & t \in [0, t_g], \\
\dot{x}(t) &= f^2(x(t), u), & t \in [t_g, t_f], \\
x(0) &= x_0, x(t_g) = x(t_g^+).
\end{cases}
\end{align*}
\]

(8)

III. PROPERTIES OF THE TWO-STAGE SYSTEM AND IDENTIFICATION MODEL

To identify the parameters in (8), it is necessary to guarantee the existence, uniqueness and Lipschitz continuity with respect to \( u \) of the solution to (8). In this section, we prove some properties of (8) and its solution. Then a parameter identification model by a least squares technique is presented. The identifiability of the above model is also proved.

A. Properties of the two-stage system

In view of the mechanism of bio-dissimilation of glycerol to 1,3-PD, the following assumptions are hold.

\( A_1 \): The concentrations of biomass, glycerol and three products are nonnegative during the process of the batch culture, that is, \( x_i \geq 0, \forall i \in I_5 \);

\( A_2 \): The concentrations of biomass, glycerol and products can not exceed their critical values, namely, \( x_i \leq x_i^\text{m}, i \in I_5 \).

Based on the above assumptions, we will consider the properties of (8) on \( S_0 \) defined as

\[
S_0 := \{ x \in \mathbb{R}^5 | x_1 \in [x_{1s}, x_{1u}], x_2 \in [x_{2s}, x_{2u}], \\
x_i \in [0, x_i^\text{m}], i = 3, 4, 5 \},
\]

(9)

where \( x_{1s} := 0.001x_1^\text{m}, x_{2s} := 0.001x_2^\text{m} \).

Property 1: Under the assumptions (\( A_1 \)) and (\( A_2 \)), the functions \( f^i : S_0 \times U_{\text{ad}} \rightarrow \mathbb{R}^5, i = 1, 2, \) defined by (6) satisfy that there exists an \( L < \infty \), such that, for any \( x, y \in S_0 \), and \( u, v \in U_{\text{ad}} \),

\[
\|f^i(x, u) - f^i(y, v)\| \leq L(\|x - y\| + \|u - v\|),
\]

(10)

where \( \| \cdot \| \) denotes the Euclidean norm.
Proof. For any \( x, y \in S_0 \) and \( u, v \in U_{ad} \), we denote \( y = x + \Delta x, v = u + \Delta u \). Then, by the differential mean value inequality, we obtain that
\[
\| f^i(y, v) - f^i(x, u) \| \\
\leq \| \frac{\partial f^i}{\partial x} (x + \theta_1 \Delta x, u + \Delta u) \| \| \Delta x \| + \\
\| \frac{\partial f^i}{\partial u} (x, u + \theta_2 \Delta u) \| \| \Delta u \|,
\]
where \( 0 < \theta_1, \theta_2 < 1 \). In view of (6), let
\[
A_{i1} := \frac{\partial f^i}{\partial x} (x + \theta_1 \Delta x, u + \Delta u), \\
A_{i2} := \frac{\partial f^i}{\partial u} (x, u + \theta_2 \Delta u).
\]
It is easy to show that \( \| A_{ij} \|, i = 1, 2, j = 1, 2 \), are bounded. Consequently, there exist \( L_{ij} > 0 \) such that \( \| A_{ij} \| \leq L_{ij} \). Take \( L = \max \{ L_{ij} \} \), then it follows that (10) holds. □

**Corollary 3.1:** Under the assumptions (A1) and (A2), \( f^i, i = 1, 2 \), satisfy the linear growth condition, i.e., there exists a nonnegative constant \( K > 0 \) such that \( \| f^i(x, u) \| \leq K(\| x \| + 1) \).

**Proof.** Take \( y = \frac{x}{2}, u = v \) to (10), then
\[
\| f^i(x, u) \| - \| f^i(\frac{x}{2}, u) \| \leq \| f^i(\frac{x}{2}, u) - f^i(x, u) \| \leq \frac{L}{2} \| x \|.
\]
Since \( U_{ad} \) is a compact set and \( f^i(\frac{x}{2}, u) \) is Lipschitz continuous in \( u \) on \( U_{ad} \), \( f^i(\frac{x}{2}, u) \) is bounded on \( U_{ad} \). That is, there exists a \( K_0 \) such that \( \| f^i(\frac{x}{2}, u) \| \leq K_0 \). Take \( K = \max \{ K_0, \frac{L}{2} \} \), which completes our proof. □

**Property 2:** Under the assumptions (A1) and (A2), the system (8) has a unique solution, \( x(t; u) \), for any \( u \in U_{ad} \) and \( t \in [0, t_f] \). Moreover,
\[
x(t; u) = \begin{cases} x_0 + \int_0^t f^1(x(s; u), u)ds, & t \in [0, t_g), \\
x(t_g; u) + \int_{t_g}^t f^2(x(s; u), u)ds, & t \in (t_g, t_f]. 
\end{cases}
\]

(11)

Furthermore, \( x(t; u) \) is Lipschitz continuous in \( u \) on \( U_{ad} \).

**Proof.** Since \( f^1(x(t), u) \) is Lipschitz continuous with respect to \( x \) in \( S_0 \) from (6), the system (8) has a unique solution by the theory of differential equations. Now, we prove that \( x(t; u) \) is Lipschitz continuous in \( u \) on \( U_{ad} \).

**Case 1.** The case for \( t \in [0, t_g] \).

According to (10) and (11), we obtain that, for any \( u, v \in U_{ad} \),
\[
\| x(t; u) - x(t; v) \| \\
\leq L t_g \| u - v \| + L \int_0^t \| x(s; u) - x(s; v) \| ds.
\]

By Gronwall inequality, we have
\[
\| x(t; u) - x(t; v) \| \leq L t_g \| u - v \| \exp(L t_g).
\]

**Case 2.** The case for \( t \in (t_g, t_f] \).

By the similar proof of case 1, we conclude that
\[
\| x(t; u) - x(t; v) \| \\
\leq L (t_g \exp(L t_g) + (t_f - t_g)) \exp(L(t_f - t_g)) \| u - v \|.
\]

Based on the above two cases, the Lipschitzian continuity is concluded. □

**Property 3:** Under the assumptions (A1) and (A2), the solution (11) is bounded, that is, there exists an \( M > 0 \) such that, for any \( t \in [0, t_f] \),
\[
\| x(t; u) \| \leq M, \forall u \in U_{ad}.
\]

**Proof.** This conclusion can be drawn from Corollary 1 and Gronwall inequality. □

Given \( x_0 \in S_0 \), we define the set of solutions to the system (8) as
\[
S_1(x_0) := \{ x(t; u) \in R^5 | x(t; u) \text{ is a solution to the system (8) for all } u \in U_{ad} \}.
\]

(13)

In view of the compactness of \( U_{ad} \subseteq R^{18} \) and the continuity of the mapping from \( u \in U_{ad} \) to \( x(t; u) \), we have the following result.

**Property 4:** The set \( S_1(x_0) \) defined by (13) is compact in \( C^1([0, t_f]; R^5) \).

**B. Identification model**

Given \( x_0 \), we have observed \( l \) experimental data in batch culture. The concentrations of biomass, glycerol, 1,3-PD, acetic acid and ethanol measured at time \( t_j \) in the experiment are denoted by \( y_{j1}, \ldots, y_{j5}, j \in I_1 := \{ 1, \ldots, l \} \), respectively, where \( t_j \) is the moment of recording experimental data. Let \( y_j := (y_{j1}, \ldots, y_{j5})^T \in R^5 \). Furthermore, take the errors between the computational values and experimental data as the cost function defined by
\[
J(u) := \sum_{j=1}^l \| x(t_j; u) - y_j \|^2.
\]

(14)

Where \( x(t_j; u) \) is the solution to the system (8) at time \( t_j \). Hence, the parameter identification model (PIP) can be formulated as
\[
(PIP) \quad \min_{u \in U_{ad}} J(u) \quad \text{s.t.} \quad x(t; u) \in S_1(x_0), t \in [0, t_f].
\]

(15)

**Theorem 1:** Under the assumptions (A1) and (A2), there exists at least an optimal solution of (PIP), i.e., there exists a \( u^* \in U_{ad} \) such that \( J(u^*) \leq J(u), \forall u \in U_{ad} \).

**Proof.** Let the feasible region \( C_0 \) of (PIP) be
\[
C_0 := \{ u \in U_{ad} | x(t; u) \in S_1(x_0), \forall t \in [0, t_f] \}.
\]

(16)

In view of the assumptions (A1) and (A2), it’s known that \( C_0 \) is non-empty. Since \( U_{ad} \) is a compact set in \( R^{18} \) and \( C_0 \) is a subset of \( U_{ad} \), we can conclude that \( C_0 \) is bounded. Furthermore, \( C_0 \) is closed by its definition (16). Hence, \( C_0 \) is compact in \( R^{18} \). In addition, \( J(u) \) is continuous in \( u \) on \( C_0 \) by the property 2, so there exists at least an optimal solution of (PIP). □
IV. OPTIMIZATION ALGORITHM AND NUMERICAL RESULTS

A. Optimization algorithm

Simulated annealing firstly appeared in 1953 [10]. The basic concept of simulated annealing originates from the physical process of metallurgical annealing. An annealing process occurs when a metal in a heat bath is initially at high temperature and is slowly cooled. At first, all the particles are distributed randomly in a quasi-liquid state. As the temperature drops, particles arrange themselves in a low-energy ground state (i.e., at or very close to the global minimum of energy), forming a crystal. Model parameters of an optimization problem play the role of particles in an idealized physical system. The objective function is an analogue of the energy function, which is to be minimized[11]. Recently, many researchers have made some improvements on the original algorithm, such as [12] and [13] etc.

In this section, we put forward an improved simulated annealing(ISA) for solving (PIP) quickly. In the algorithm, the local search algorithm (Hooke-Jeeves) is embedded into and only one component of the current solution is changed in each iteration. In addition, in order to reduce the computation cost, we adjust the iterative times at certain temperature according to the annealing temperature. The main steps of ISA are as follows:

Step 1 Initialize the highest temperature $T_{\text{max}}$, the lowest temperature $T_{\text{min}}$, the largest inner iteration times $L_{1\text{max}}, L_{2\text{max}}$, the temperature reducing constant $dt$ and the admissible parameter region $U_{\text{ad}}$. Set the annealing temperature $T := T_{\text{max}}$ and the inner iteration times $k = 0$.

Step 2 Randomly generate an initial feasible solution $u^{0} \in U_{\text{ad}}$. Solve the system (8) and obtain $x(t; u^{0})$ by Runge-Kutta method. Then calculate the fitness value $J(u^{0})$ by the formula (15). Set the optimal parameter $u^{*} := u^{0}$, $u := u^{0}$ and the optimal fitness value $J^{*} = J(u^{*})$.

Step 3 While $T > T_{\text{min}}$

If $T > T_{0}$ (a given temperature constant), $L_{\text{max}} = L_{1\text{max}}$; otherwise $L_{\text{max}} = L_{2\text{max}}$.

While $k < L_{\text{max}}$

3.1 Randomly select an integer $r$ from the set $I_{18}$, let $u_{r}' = u_{r} + \alpha(u^{0}_{r} - u_{r})\xi_{r}, \xi_{r} \sim N(0, 1)$, where $N(0, 1)$ denotes a standard normal distribution and $\alpha$ is a variable which decreases by the formula $\alpha = \alpha e^{-\beta}$. If $\alpha < 10^{-5}$, then $\alpha = 1$. Readjust the component $u_{r}'$ by (17), and the other components of $u^{0}$ are unchanging. So the new solution generated is denoted by $u'$, $u' = (u_{1}, \ldots, u'_{r}, \ldots, u_{18})^T$.

$$u_{r}' = \begin{cases} u_{r}^{0} - (u_{r}^{0} - u_{r}'), & u_{r}' < u_{r}^{0} \\ u_{r}', & u_{r}^{0} < u_{r}' < u_{r}^{0} \\ u_{r}^{0} + (u_{r}^{0} - u_{r}'), & u_{r}' > u_{r}^{0} \end{cases}$$ (17)

3.2 If a random number $p < P_{0}(0 < P_{0} < 1)$, then Hooke-Jeeves search will be carried on and the found local optimal solution is denoted by $u'$.

3.3 Evaluate the fitness value $J(u')$ and the change in energy level $\Delta J = J(u') - J(u)$. If $\Delta J \leq 0$, then accept the new solution; else accept the new solution with a certain probability $p = \exp(-\frac{\Delta J}{T})$. If the new solution $u'$ is accepted, then set $u = u'$ and $k = k + 1$.

3.4 If $J(u) < J(u')$, then $u^{*} = u$. Reduce the temperature $T$ by $T = T \ast dt$.

Step 4 Output the optimal parameter $u^{*}$ and the optimal fitness value $J^{*}$.

B. Numerical results

According to the 7 groups of experimental data, we obtain the optimal parameter $u^{*}$ applying the above algorithm ISA. In ISA, the parameters $T_{\text{max}} = 5; L_{1\text{max}} = 10; L_{2\text{max}} = 30; dt = 0.8; \beta = 1.01; P_{0} = 0.6$; the initial value of $\alpha$ is equal to 1. The optimal parameters and the relative errors $e_{1i}$ and $e_{2i}$ between the computational values and the experimental data are listed in Table 1, where $e_{1i}$ denotes the errors in this paper and $e_{2i}$ represents the ones in [6]. The relative errors in Table 1 are defined as follows:

$$e_{1i} = \frac{\left( \sum_{i=1}^{I} [x(t; u^{*}) - y_{ji}]^2 \right)^{\frac{1}{2}}}{\left( \sum_{i=1}^{I} [y_{ji}]^2 \right)^{\frac{1}{2}}}, i \in I_{5}, \quad (18)$$

$$e_{2i} = \frac{\left( \sum_{i=1}^{I} [x(t; u_{0}) - y_{ji}]^2 \right)^{\frac{1}{2}}}{\left( \sum_{i=1}^{I} [y_{ji}]^2 \right)^{\frac{1}{2}}}, i \in I_{5}, \quad (19)$$

where $x(t; \cdot)$ is the computational value of the concentration at $t_{g}$, $u_{0}$ is the optimal parameter found in [6]. From the Table 1, we conclude that the relative errors between the computational values and the experimental data are reduced greatly.

Finally, we simulate the process of the batch fermentation according to the optimal parameter $u^{*}$. Fig. 1 shows the comparison of concentrations of glycerol, 1,3-PD, acetic acid and ethanol between computational results on $u^{*}$ and the experimental results, respectively. Fig. 2 shows the comparison of the concentration of biomass between computational results on $u^{*}$ and the experimental results, respectively. In these figures, the curves represent the computational results and the points denote the experimental data. From Fig. 1 and Fig. 2, we can conclude that the two-stage dynamical system describes well not only the developmental and growth phases but also the stationary phase. Numerical results also show that the optimization algorithm is valid.

V. CONCLUSIONS AND FUTURE WORKS

In this paper, we propose a two-stage nonlinear dynamical system to describe the process of batch fermentation. In the
above numerical example, the comparison of the errors between in this paper and in [6] show that the two-stage system is more suitable for the batch process. Errors between the computational and experimental results are reduced greatly, which is helpful to microbial production of 1,3-PD. In addition, we prove some prime properties of the two-stage system and establish a parameter identification model. In the future work, we will investigate the optimal control model to enhance the productivity of 1,3-PD by optimizing initial condition and terminal time in the batch culture. Furthermore, the optimality conditions and optimization algorithm of the optimal control model will be also studied.

REFERENCES
