IDENTIFICATION OF YIELD COEFFICIENTS IN A BAKER'S YEAST MODEL AN OPTIMAL EXPERIMENTAL DESIGN APPROACH

C. ROCHA and E.C. FERREIRA
Departamento de Engenharia Biológica
Universidade do Minho
Campus de Gualtar
4710 Braga Portugal

e-mail: ecferreira@deb.uminho.pt
http://www.deb.uminho.pt

ABSTRACT

The main objective of this work is to elaborate methodologies that allow the identification of yield coefficients through complete measurements of the state. Experimental design strategies are proposed in order to optimize the richness of data coming out from the experiments, quantified by indexes related to the Fisher information matrix. The objectives of the experimental planning have been addressed in terms of the programming of input trajectories. The experimental planning is envisaged for baker's yeast aiming at the computation of the substrate feed trajectories.

KEYWORDS

Baker's Yeast, Parameter Estimation, Yield Coefficients, Experiment Design, Optimization

INTRODUCTION

The dynamics of a reactor network in a stirred tank bioreactor can be described by the following mass balance equations written in the matrix form:

$$\frac{d\xi}{dt} = Kr(\xi) - D\xi + U \tag{1}$$

where ξ is the vector of the n component concentrations, K is the matrix of yield coefficients, r is the vector of reaction rates, U is the vector of feed rates and of gaseous outflow rates, D is the dilution rate. The reaction network for baker's yeast (*Saccharomyces cerevisiae*) growth on sugar with ethanol production /consumption is usually [1] described by the following three metabolic pathways:

$$\textit{Respiratory growth on sugar}: \quad k_1S + k_5C \xrightarrow{r_1} X + k_7P$$

$$\textit{Fermentative growth on sugar}: \quad k_2S \xrightarrow{r_2} X + k_8P + k_3E$$

$$\textit{Respiratory growth on ethanol}: \quad k_4E + k_6C \xrightarrow{r_3} X + k_9P$$

where S, C, X, P, E represent sugar (glucose), oxygen, yeast, carbon dioxide, and ethanol respectively, r_1, r_2, r_3, are the reaction rates. These reactions are autocatalytic since yeast are self-reproducing microorganisms. The yield coefficients k_i are expressed with respect to the production of 1 unit of yeast in each reaction.

Ethanol can not be simultaneously produced and consumed. Glucose and ethanol can be cometabolized as long as the respiratory capacity is not exceeded (subcritical glucose flux). The correspondent metabolic state is designated oxidative or respiratory. Under conditions of supracritical glucose flux, ethanol is produced - oxidoreductive or respiro-fermentative pathways. The associated partial dynamical models are represented as follows [2]

$$\frac{d}{dt}\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -k_1 & -k_2 \\ 0 & k_3 \\ -k_5 & 0 \\ k_7 & k_8 \end{bmatrix} \begin{bmatrix} r_1(\xi) \\ r_2(\xi) \end{bmatrix} - D\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{in} \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad (2a) \qquad \frac{d}{dt}\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -k_1 & 0 \\ 0 & -k_4 \\ -k_5 & -k_6 \\ k_7 & k_9 \end{bmatrix} \begin{bmatrix} r_1(\xi) \\ r_3(\xi) \end{bmatrix} - D\begin{bmatrix} X \\ S \\ E \\ O \\ C \end{bmatrix} + \begin{bmatrix} 0 \\ DS_{in} \\ 0 \\ OTR \\ -CTR \end{bmatrix} \quad (2b)$$

The identification problem in bioprocess dynamical modelling is usually twofold: the estimation of the yield coefficients k_i in matrix K; and the determination of a suitable structure for the reaction rate model $r(\xi)$ and the estimation of the kinetic coefficients involved in $r(\xi)$. Only the former case will be addressed in this work following an approach introduced by Chen [3].

Defining a nonsingular partition of the matrix K in two parts K_a, K_b such that K_a has full rank and introducing the following linear state transformation:

$$Z = C\xi_a + \xi_b \quad (3)$$

where the (n-p)xp matrix C is the solution of the matrix equation

$$CK_a + K_b = 0 \quad (4)$$

The initial model (1) is then rewritten as

$$\frac{d\xi_a}{dt} = K_a r - D\xi_a + U_a \quad (5a)$$

$$\frac{dZ}{dt} = -DZ + CU_a + U_b \quad (5b)$$

The important point is that the auxiliary model (5b) is independent of the reaction rates. Z can be decomposed as follows

$$Z = CZ_a + Z_b \quad (6)$$

where the variables Z_a and Z_b are governed by the following dynamics:

$$\frac{dZ_a}{dt} = -DZ_a + U_a \quad (7a)$$

$$\frac{dZ_b}{dt} = -DZ_b + U_b \quad (7b)$$

Substituting (6) in (3), ξ_b is computed as

$$\xi_b = Z_b - C(\xi_a - Z_a) \quad (7c)$$

If we set $y = Z_b - \xi_b$ and $\phi = \xi_a - Z_a$ the equation (7c) take the following standard linear regression form

$$y(t) = C\phi(t) \quad (8)$$

In the case of the bakers yeast , the following regression equations are obtained

$$\begin{bmatrix} X - Z_{b1} \\ C - Z_{b2} \\ E - Z_{b3} \end{bmatrix} = \begin{bmatrix} \theta_1 & \theta_2 \\ \theta_3 & \theta_4 \\ \theta_5 & \theta_6 \end{bmatrix} \begin{bmatrix} Z_{a1} - S \\ Z_{a2} - O \end{bmatrix} \quad (9)$$

The dynamics of Z is computed as

$$\frac{d}{dt}\begin{bmatrix} Z_{a1} \\ Z_{a2} \\ Z_{b1} \\ Z_{b2} \\ Z_{b3} \end{bmatrix} = -D\begin{bmatrix} Z_{a1} \\ Z_{a2} \\ Z_{b1} \\ Z_{b2} \\ Z_{b3} \end{bmatrix} + \begin{bmatrix} DS_{in} \\ OTR \\ 0 \\ -CTR \\ 0 \end{bmatrix} \tag{10}$$

Integration of eq. (10) allows the fitting of the regression equations (9). Yield coefficients (each partial model shares 3 coefficients) are finally recovered from the values of θ_t.

To compose a basis for the comparison of different experimental conditions with respect to parameter identification a measure of the preciseness of the estimated parameters is required. This depends on the degree of richness or degree of independence of the regressor. A measure of the accuracy of estimation is the parameter covariance matrix. The asymptotic covariance matrix of any unbiased estimator $\hat{\theta}$ of θ satisfies the inequality (also known as the Cramér-Rao lower bound): $cov\left(\hat{\theta}\right) \geq F_I^{-1}$ where F_I is the *Fisher Information Matrix* (FIM) given by $F_I = \sum_i \phi^T\left(t_i\right)\phi\left(t_i\right)$.

The D-optimality criterion of the degree of richness or degree of independence in the optimal search of experimental conditions is considered. *D-optimality design* corresponds to maximizing the determinant of FIM and is equivalent to minimising the volume of the deviation ellipsoids, a measure of the absolute errors of the estimated value of parameters.

To find the input functions that lead to the most informative experiment, the feed rate was discretized into a finite number of node points and considered to be linear between node points. For the optimization of the input variables at node points, a simplex-simulated annealing approach [4] and an adaptive random search method [5] have been used.

RESULTS AND DISCUSSION

The optimization was made under the following general conditions: glucose concentration in the feeding medium - 15 g/L; initial volume - 2 L; final volume - 5 L; operation time - 20 h. The kinetic model employed is the implementation of the “bottleneck principle” of the respiratory capacity of the yeast proposed by Sonnleitner and Käpelli [1].

The feed rate profile (fig.1) was obtained by maximization of the determinant of the FIM. As mentioned previously, there are two partial dynamic models and it was necessary to evaluate two FIM and two determinants. If the determinants had been maximized independently, only six yield coefficients could be identified for each optimization. In order to identify all the yield coefficients, the product of the two determinants was maximized.

The state variable profiles simulated with the optimized feed profile are shown in fig. 2 and 3. White noise was introduced trying to simulate experimental noise and measurement errors.

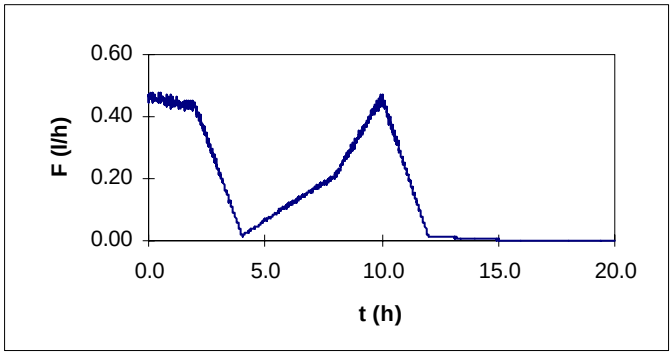

Fig. 1 Computed feed rate profile

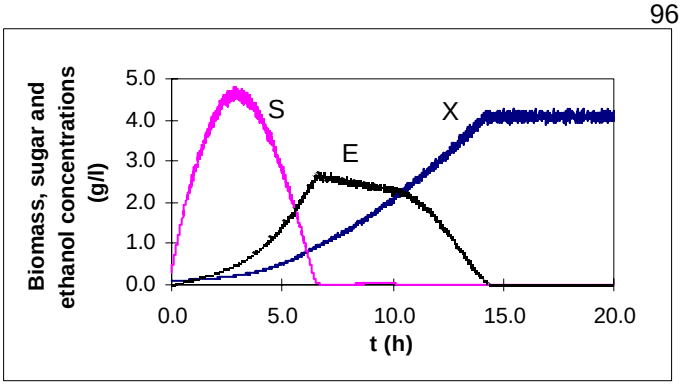

Fig. 2 Biomass, sugar and ethanol profiles

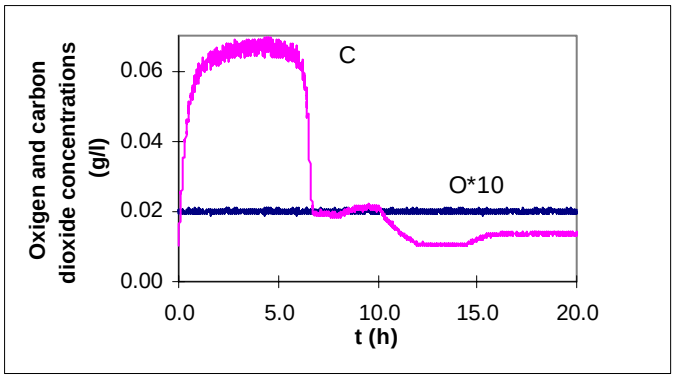

Fig. 3 Dissolved oxygen (O) and carbon dioxide (C) profiles

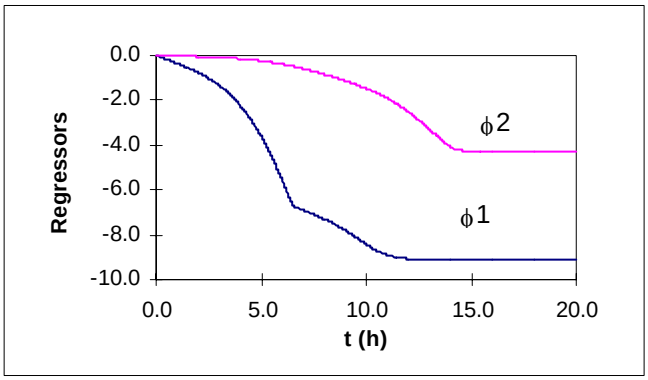

Fig. 4 Regressors for the identification experiment

Table I - Values of the identified yield coefficients

case	k_1	k_2	k_3	k_4	k_5	k_6	k_7	k_8	k_9
'real'	2.04	20.0	9.62	1.39	0.833	1.56	1.23	9.26	0.903
simulation	2.03	20.0	9.61	1.37	0.834	1.52	1.23	9.32	0.865

It results from fig. 4 that the two regressors are independent and the parameters are thus identifiable.

One can observe from Table 1 that differences between the identified parameter values and the parameter values used in simulation (from [1]) are not significant. Thus the optimal input signal suggested leads to a "rich" identification experiment.

In this study, the dissolved oxygen concentration was maintained at a constant value. It would be interesting, in the future, to optimize also the dissolved oxygen concentration, manipulating the gaseous inflow rate and agitation rate, in order to improve results.

CONCLUSIONS

For the biotechnological process under consideration, it was be shown that the yield coefficients can be reasonably identified using the data from a fed-batch experiment with an optimized feed flow pattern. Experiments at lab scale are in progress to evaluate and validate this approach.

REFERENCES

[1] Sonnleitner, Käpelli, *Biotechnol. Bioeng*. 28, 927-937, 1986.
[2] Ferreira, E.C. *Ph.D. dissertation* (in portuguese), Porto, 1995.
[3] Chen, L. *Ph.D. dissertation*, Louvain-la-Neuve, 1992.
[4] Cardoso, M.F., Salcedo, R., Azevedo, S.F. *Ind. Eng. Chem. Res.* 33:8, 1908-1918, 1994.
[5] Salcedo, R. *Ind. Eng. Chem. Res.* 31:1, 262-273, 1992.